Sanjeev Gupta · Nagasamy Nadarajan
Debjyoti Sen Gupta *Editors*

Legumes in the Omic Era

Springer

Legumes in the Omic Era

Sanjeev Gupta • Nagasamy Nadarajan
Debjyoti Sen Gupta
Editors

Legumes in the Omic Era

Springer

Editors
Sanjeev Gupta
Indian Institute of Pulses Research
Kanpur, UP, India

Nagasamy Nadarajan
Indian Institute of Pulses Research
Kanpur, UP, India

Debjyoti Sen Gupta
Department of Plant Sciences
North Dakota State University
Fargo, ND, USA

ISBN 978-1-4614-8369-4 ISBN 978-1-4614-8370-0 (eBook)
DOI 10.1007/978-1-4614-8370-0
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013951148

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Enormous amount of biological information are available today, particularly after the completion of whole genome sequencing project in legumes like *Medicago truncatula*, Soybean, Pigeonpea and Chickpea. Large-scale sequencing projects on two other legumes, *Lotus japonicus* and *Vigna radiata*, are also near completion. The information generated from genome sequencing calls for producing a complete functional interpretation of whole genome. This demand coincides with another technological development in plant biology called "OMICS" revolution. The technologies such as transcriptomics, proteomics and metabolomics are being developed in major legume species with the aim to analyze molecular data on a genome-wide scale. These developments are now becoming major landmarks in understanding legume biology in a precise manner.

The present book is an excellent review of the recent advances in the grain legumes' genomics research and applications. In this book efforts have been made to gather and present available recent information for individual grain legume species in a logical order. Genomic resources, structural and functional genomics, progress towards whole genome sequencing and use of genome sequence information in crop improvement are major aspects which are described in detail for each grain legume species in respective chapters. More focus is given to showcase the potential and practical use of genomic tools and resources available today in these species for crop improvement. Information is also shared on the advances in bioinformatics tools and techniques in grain legumes research. The genomic tools' used in revealing legume genome evolution are also discussed in detail. Legume biofortification research and importance of genomic tools in nutritional improvement of grain legumes are presented briefly.

This book contains 15 chapters authored by scientists/researchers who are actively involved in analyzing and improving particular legume genome. Their contribution is enormous in presenting up-to-date information on the subject. Some figures included by the authors in the respective chapter were published elsewhere previously. The necessary permission has been obtained by the authors to use them again for their chapters. We record our acknowledgements to all such publishers and authors for their generosity and goodwill. There are many people around the globe

who were there always during the entire developmental period of this book influencing positively to make this project feasible for us: Dr. Shiv Kumar, ICARDA, Rabat, Morocco, Dr. Rajeev Varshney, ICRISAT, Patancheru, India, Dr. Clare J Coyne, USDA/ARS, WSU, Pullman, USA, and Dr. Dil Thavarajah, NDSU, Fargo, USA.

We editors are highly thankful to Dr. S. Ayyappan, Director General, ICAR, and Secretary, DARE, Ministry of Agriculture, Govt. of India for being the source of guidance and encouragement to take up this project. Thanks are also due to Dr. S. K. Datta (DDG, Crop Science, ICAR) and Dr. B. B. Singh (ADG, Oilseeds and Pulses, ICAR) for continuous support and encouragement. We thank our parent organization, Indian Council of Agricultural Research (ICAR) for supporting the scientific endeavor like "Legumes in the Omic Era".

We thank our families for being patient and supportive in this long journey, without their moral support it would not be possible. We thank Misses Hannah Smith and Melissa Higgs (Associate Editors, Springer Science+Business Media, New York) and Mr. Michael D. Sova (Developmental Editor) for enormous support to make the publication a reality. Mr. Ashutosh Kumar, Senior Research Fellow, Crop Improvement Division of Indian Institute of Pulses Research, Kanpur, is highly appreciated for typographical help.

Kanpur, UP, India — Sanjeev Gupta
Kanpur, UP, India — Nagasamy Nadarajan
Fargo, ND, USA — Debjyoti Sen Gupta

Contents

Contributors

Tuba K. Anjum, M.Sc. Crop Improvement Division, Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Francisco J.L. Aragao, Ph.D. Embrapa Recursos Genéticos e Biotecnologia, Norte Brasília, DF, Brazil

Michael Baum, Ph.D. Biodiversity and Integrated Gene Management (BIGM) Program, Center for Agricultural Research in the Dry Areas (ICARDA), Aleppo, Syria

Abhishek Bohra, Ph.D. Crop Improvement Division, Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

P. Dhanasekar, M.Sc. (Ag.) Nuclear Agriculture and Biotechnology Division, Bhabha Atomic Research Centre, Mumbai, Maharashtra, India

Rebecca Dickstein, Ph.D. Department of Biological Sciences, University of North Texas, Denton, TX, USA

Pooran M. Gaur, Ph.D. Grain Legume Genetics Program, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Andhra Pradesh, India

Debjyoti Sen Gupta, M.Sc. (Ag.) Cereal Science Graduate Program, Department of Plant Sciences, North Dakota State University, Fargo, ND, USA

Sanjay Gupta, Ph.D. Crop Improvement Section, Directorate of Soybean Research, Khandwa Road, Indore, MP, India

Sanjeev Gupta, Ph.D. Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Aladdin Hamwieh, Ph.D. Biodiversity and Integrated Gene Management (BIGM) Program, International Center for Agricultural Research in the Dry Areas (ICARDA), Aleppo, Syria

S.M. Husain, Ph.D. Crop Improvement Section, Directorate of Soybean Research, Indore, MP, India

Arvind M. Kayastha, Ph.D. Faculty of Sciences, School of Biotechnology, Banaras Hindu University, Varanasi, Uttar Pradesh, India

Eva Konečná, Ing. Department of Plant Science, Mendel University in Brno, Brno, Czech Republic

Jitendra Kumar, Ph.D. Crop Improvement Division, Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Shiv Kumar, Ph.D. Biodiversity and Integrated Gene Management (BIGM) Program, International Center for Agricultural Research in the Dry Areas, Rabat, Morocco

Giriraj Kumawat, Ph.D. Crop Improvement Section, Directorate of Soybean Research, Indore, MP, India

Marc Libault, Ph.D. Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK, USA

Nupur Malviya, Ph.D. Crop Improvement Division, Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Alagu Manickavelu, Ph.D. Biodiversity and Integrated Gene Management (BIGM) Program, International Center for Agricultural Research in the Dry Areas, Rabat, Morocco

Pooja Bhatnagar Mathur, Ph.D. International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, Andhra Pradesh, India

Phillip E. McClean, Ph.D. Department of Plant Sciences, North Dakota State University, Fargo, ND, USA

Fred J. Muehlbauer, Ph.D. Grain Legume Genetics Physiology Research Unit, USDA, WSU, Pullman, WA, USA

Nagasamy Nadarajan, Ph.D. Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Juan M. Osorno, Ph.D. Department of Plant Sciences, North Dakota State University, Fargo, ND, USA

Aditya Pratap, Ph.D. Crop Improvement Division, Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Neha Rajan, M.Sc. Krishi Vigyan Kendra, Ramakrishna Mission Ashram, Ranchi, Jharkhand, India

S.V. Ramesh, Ph.D. Crop Improvement Section, Directorate of Soybean Research, Indore, MP, India

Milind B. Ratnaparkhe, Ph.D. Crop Improvement Section, Directorate of Soybean Research, Indore, MP, India

Gopesh C. Saha, Ph.D. Department of Crop and Soil Sciences, Washington State University, Pullman, WA, USA

C.V. Sameerkumar, Ph.D. Pigeonpea Breeding, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, Andhra Pradesh, India

K.B. Saxena, Ph.D. Pigeonpea Breeding, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, Andhra Pradesh, India

Rachit K. Saxena, Ph.D. Center of Excellence in Genomics (CEG), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, Andhra Pradesh, India

Tilak Raj Sharma, Ph.D. Department of Agricultural Biotechnology, CSK Himachal Pradesh Agricultural University, Palampur, Himachal Pradesh, India

A.K. Singh, Ph.D. Department of Genetics and Plant Breeding, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi, Uttar Pradesh, India

B.D. Singh, Ph.D. Faculty of Sciences, School of Biotechnology, Banaras Hindu University, Varanasi, Uttar Pradesh, India

B.B. Singh, Ph.D. Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, USA

Mritunjay Singh, M.Sc. Crop Improvement Division, Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Vinay Kumar Singh, M.Sc. Faculty of Sciences, Centre for Bioinformatics, School of Biotechnology, Banaras Hindu University, Varanasi, Uttar Pradesh, India

Petr Smýkal, Ph.D. Department of Botany, Palacky University in Olomouc, Olomouc, Czech Republic

J. Souframanien, Ph.D. Nuclear Agriculture and Biotechnology Division, Bhabha Atomic Research Centre, Mumbai, Maharashtra, India

Srinivasan Samineni, Ph.D. Grain Legume Genetics Program, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Andhra Pradesh, India

Ekta Srivastava, M.Sc. Crop Improvement Division, Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Dil Thavarajah, Ph.D. School of Food Systems, North Dakota State University, Fargo, ND, USA

Pushparajah Thavarajah, Ph.D. School of Food Systems, North Dakota State University, Fargo, ND, USA

Mahendar Thudi, Ph.D. Grain Legume Genetics Program, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Andhra Pradesh, India

Michael P. Timko, Ph.D. Department of Biology, University of Virginia, Charlottesville, VA, USA

Rakhi Tomar, M.Sc. Crop Improvement Division, Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Rajeev K. Varshney, Ph.D. Center of Excellence in Genomics (CEG), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, Andhra Pradesh, India

Chapter 1
Legumes in Omic Era: Retrospects and Prospects

Sanjeev Gupta, Nagasamy Nadarajan, and Debjyoti Sen Gupta

Abstract Legumes are important for nutritional security of humans and livestocks and ecological sustainability of agricultural production systems of the world. The adaptability and productivity of legumes are limited by major biotic and abiotic stresses. Therefore, there is a crucial need to increase tolerance against various stresses, which is a major challenge in legume improvement programs for enhancing yield. Breeding methods complemented adequately by genomics approaches could lead to simpler and more effective gene-based approach for legume improvement. This requires adequate genomic resources and information for each legume species of economic importance. Major developments made in recent past, like genome sequencing, the "omic" research and bioinformatics have provided scope for utilization of genomic resources for legume improvement. A good progress has been made in genome sequencing of some legumes and this will increase even more due to novel sequencing technologies called next generation sequencing. Since the release of genome sequences of *Lotus japonicus*, *Glycine max*, *Medicago truncatula*, *Cajanus cajan* and *Cicer arietinum*, a number of comprehensive tools such as bioinformatics tools for sequence assembly and functional annotation, microarray platforms for high-throughput gene expression, transformation systems, and large cDNA and gDNA libraries have been developed for important legumes. These tools need to be integrated to understand genome structure and function of legumes. More comprehensive approaches, including quantitative and qualitative analyses of gene expression

S. Gupta, Ph.D. (✉)
Indian Institute of Pulses Research,
Kanpur, Uttar Pradesh 208024, India
e-mail: saniipr@rediffmail.com

N. Nadarajan, Ph.D.
Indian Institute of Pulses Research,
Kanpur, Uttar Pradesh, India

D.S. Gupta, M.Sc. (Ag.)
Cereal Science Graduate Program, Department of Plant Sciences, North Dakota State University, Fargo, ND, USA

S. Gupta et al. (eds.), *Legumes in the Omic Era*, DOI 10.1007/978-1-4614-8370-0_1,

products are further necessary at the transcriptomic, proteomic, and metabolomic levels for better understanding the functioning of genomes and their gene, including their regulatory networks by combining at computational approaches with translational genomics. The progress in omics research will considerably contribute to better understanding of the molecular and genetic basis of yield and tolerance to biotic and abiotic stress and accelerate molecular and transgenic breeding of legumes of economic importance.

Keywords Legumes • Genome sequence • Transcriptomics • Proteomics • **Metabolomics**

Introduction

Legumes are the third largest family of higher plants with more than 20,000 species having major impact on agriculture, human and livestock nutrition and environment. These are second only to grasses in agricultural importance (Doyle 2001). Major grain legumes like common beans (*Phaseolus* spp.), pea (*Pisum sativum* L.), chickpea (*Cicer arietinum* L.), broad bean (*Vicia faba* L.), pigeonpea (*Cajanus cajan* L.), cowpea (*Vigna unguiculata* L.), and lentil (*Lens esculentum* L.) together constitute 33 % of dietary protein needs of human (Vance et al. 2000). Moreover, grain legumes, predominantly soybean (*Glycine max* L.) and peanut (*Arachis hypogeae* L.), are also a major source for vegetable oil, providing more than 35 % of the world's processed vegetable oil. Among fodder legumes, medics (*Medicago* spp. L.), clovers (*Trifolium* spp.), Vetches (*Vicia* spp.) and Stylos (*Stylosenthes* spp.) assume major importance for animal productivity and nitrogen economy in most parts of the world.

Most of the grain legumes are attributed for low yield. The adaptability and productivity of legumes are limited by major biotic and abiotic stresses, including fungal and viral diseases, insect pests, drought, heat, frost, chilling, water-logging, salinity and mineral toxicities (Dita et al. 2006). Fusarium wilt, blights and viral diseases inflicts severe losses. Similarly, pod borers and sucking pests also causes serious problems in crop management. Foliar and root diseases in forage legumes also constitute major production constraints. Hence, there is a crucial need to increase the abiotic and biotic stress tolerance in legumes, which is a major challenge in crop improvement programs for enhancing yield. Although conventional plant breeding and enhanced management strategies have addressed several constraints that limit crop productivity or quality, there are situations where the existing genetic resources lack the requisite traits. Breeding methods complemented adequately by genomics approaches could lead to simpler and more effective gene-based approach for legume improvement. This requires adequate genomic resources and information for each legume species of economic importance.

In the beginning, the progress in legume biology has been made by the development of model systems to investigate the genetics of important traits. Like *Arabidopsis thaliana*, the two legumes, *Lotus japonicus* and *Medicago truncatula*,

due to their small and diploid genomes, autogamous nature, short generation times, and prolific seed production were emerged out as model legume plant systems (Cook 1999; Handberg and Stougaard 1992). Since then, powerful genetic and genomic tools have been developed that include genome sequencing (Kato et al. 2003), isolation of expressed sequence tags (ESTs) (Asamizu et al. 2004; Kulikova et al. 2001), and establishment of genetic and physical maps for each model species (Pedrosa et al. 2002; Thoquet et al. 2002). The increasing wealth of genetic and genomic data and the high degree of synteny between legume genomes (Kalo et al. 2004; Stracke et al. 2004) make these two species valuable models for the molecular genetic study of various traits related to increased productivity of legumes.

Recent advances in plant genomics have moved beyond model systems to various plant species of economic importance. Since the release of genome sequences of *Arabidopsis* and rice in the past (Goff et al. 2002; Yu et al. 2002; Lin et al. 1999; Mayer et al. 1999) and of *Lotus japonicus*, *Glycine max*, *Medicago truncatula*, *Cajanus cajan* and *Cicer arietinum* recently (Sato et al. 2008; Schmutz et al. 2010; Young et al. 2011; Singh et al. 2012; Varshney et al. 2010, 2011), a number of comprehensive tools such as bioinformatics tools for sequence assembly and functional annotation, microarray platforms for high-throughput gene expression, transformation systems, and large cDNA and gDNA libraries have been developed for a range of species, including the important legumes. Now a major challenge is to integrate these various tools to better understand genome structure and function.

While sequence information is invaluable and a necessary starting point, it is insufficient to answer questions concerning gene function, regulatory networks, and the biochemical pathways activated in response to biotic and abiotic stresses. To address these questions, more comprehensive approaches, including quantitative and qualitative analyses of gene expression products are necessary at the transcriptomic, proteomic, and metabolomic levels. These developments will provide opportunities for better understanding the functioning of genomes and their gene, including their regulatory networks by combining at computational approaches with translational genomics.

Retrospects

Three major developments made in recent past, like genome sequencing, the OMICS research and bioinformatics have revolutionized the plant biology (Weckwerth 2011a, b). A good progress has been made in genome sequencing of some legumes and this will increase even more due to novel sequencing technologies called next generation sequencing (Weckwerth 2011a, b; Ideker et al. 2001). Based on sequence information, genome assembly is developed. After the assembly of a full genome, functional annotation is established for each sequence assembly. Predicted genes are searched for homology against databases of characterized genes and proteins. It is obvious that this initial functional annotation is not capable of producing a complete functional interpretation of the whole genome and a prediction of the molecular phenotype (Weckwerth 2011a, b). Consequently, the molecular

phenotype needs to be measured for the functional interpretation of the genotype. This requires the support of "omics" research. A modest beginning has been made in this area for legumes. Recent progress indicates that appropriate bioinformatics platform has also been developed for utilization of sequence and "omic" information for improvement of legumes.

Genome Sequencing

The nuclear genomes of legumes vary greatly in size, from 370 Mbp in *Lablab niger* to as large as 13,000 Mbp of *Vicia faba*. Efforts have been made to generate complete sequence information of some important legumes. The whole genome sequencing of three legumes viz., *Medicago*, *Lotus* and *Glycine* has been completed till date (Young et al. 2011; Sato et al. 2008; Schmutz et al. 2010). The draft genome sequences of *Cajanus* and *Cicer* are already available (Singh et al. 2012; Varshney et al. 2010, 2011). Genome sequencing in some other legumes like mungbean, peas, alfalfa, peanut, cowpea and common bean are at various stages of progress, with the later being expected to be completed shortly. With the advent of next generation sequencing, the task of whole genome sequencing of these crops can be efficiently completed. The next-generation sequencing not only is a dramatic advance over capillary-based sequencing but also meets significant challenges in assembly and sequence accuracy due to short read lengths, method-specific sequencing errors, and the absence of physical clones. However, the promise of much lower sequencing cost with the now proven concept of next-generation expressed sequence tag sequencing which will allow assessment of plant genomes at least at the functional level (Ohtsu et al. 2007).

"OMICS" Research

"Omics" research involves functional genomics, transcriptome profiling, proteomics, and metabolomics for analyzing molecular data of living systems on a genome scale (Somerville and Dangl 2000; Weckwerth 2003; Ideker et al. 2001). This provides genome-scale molecular data in combination with a genomic template. The ultimate goal is to derive a model of metabolism that is driven by genome data and predicts the phenotype (Weckwerth 2011a, b). Transcriptomics, proteomics and metabolomics data can be exploited for gene prediction and functional gene annotation in fully sequenced organisms (Castellana et al. 2008; Wienkoop et al. 2010). Major studies in plant model systems such as *Arabidopsis thaliana* and *Chlamydomonas reinhardtii* have demonstrated the applications of proteo- and metaproteogenomics (Castellana et al. 2008; Wienkoop et al. 2010). The progress in omics research will considerably contribute to better understanding of the molecular and genetic basis of yield and tolerance to biotic and abiotic stress that has been an important bottleneck for legume molecular and transgenic breeding.

Functional Genomics

Since the 1990s, genomics has been the most active research field in biological science generating a huge amount of information, while structural genomics has emerged at the methodological level to understand gene expression and function. A broad range of genomic resources has been developed to accelerate legume improvement. These include expression sequence tag (EST) database, genome sequences (whole or partial), physical maps, molecular maps, DNA chips and bacterial artificial chromosome (BAC) or similar genomic libraries. BAC libraries have been widely used in different aspects of genome research. The value of large insert libraries has long been recognized in genome analysis. BAC libraries are important genomic resources that have been used for (1) physical mapping of genomes, (2) cost effective molecular marker development of microsatellite markers (Shultz et al. 2007), (3) map-based cloning of genes or QTL for important agronomic traits, (4) evolutionary study of multigene families, (5) karyotyping genomes through BAC-FISH, and (6) whole genome sequencing. BAC libraries have been constructed for many species and are usually developed by cloning size-fractionated DNA fragments partially digested with restriction enzymes. In near future, the BACs of legumes should have potential applications in comparative genomics and functional genomics as well owing to the macro- and microsynteny widespread within legumes. Among the grain legumes, soybean has been more intensively studied and according to the legume information system data, over 1.3 million ESTs were developed from different cDNA libraries, which is the largest in number among the individual grain legume ESTs. The availability of a large number of EST and BAC sequences facilitated the discovery of new SNP and SSR markers toward the construction of high-resolution genetic maps of various legumes. With the availability of whole genome sequence information, large numbers of ESTs are identified for biotic and abiotic response in two other legumes, chickpea and pigeonpea. However, progress made in this area is also satisfactory in cowpea and groundnut.

Transcriptomics

Unlike genome analysis, transcriptome analysis offers a full profile of gene function information under various conditions, and it differs with dissimilar environments, cell types, developmental stages, and cell states (Moe et al. 2011). It provides a powerful tool for differential gene expression, mutant splicing, SSR or SNP analysis, and functional genetics studies. The typical analysis of the dynamic transcriptome is usually performed with microarray technology and is one of the pioneering genome-scale, hypothesis-free screening methods. Several large-scale studies have revealed differential gene expression under different conditions and almost every gene in *Arabidopsis thaliana* is already characterized based on RNA sequencing data under specific conditions (www.arabidopsis.org). Nowadays, NGS provides an alternative technology for RNA sequencing (Brautigam and Gowik 2010; Wang et al. 2010). However, this technology is still in development and very expensive

because several fold genome coverage has to be measured to obtain statistically significant data. With development of new technologies it is expected that transcriptomic analysis will be performed extensively in major legumes in near future.

Since trancriptomics in legumes started its beginning, it is desirable that gene expression studies performed in *Arabidopsis*, and resulting knowledge from such studies be used in legumes through comparative genomics. For example, Ishitani et al. (2004) selected 100–200 genes from the *Arabidopsis* database and showed that at least three DREB-like genes, thought to be key transcriptional regulators of drought and/or cold tolerance, in common bean. Similarly, in Arabidopsis, analysis of the transcriptome changes occurring during cold, drought, and salt stress in a survey of 7,000 genes showed a shared response for a majority of cold and drought stress regulated genes (Seki et al. 2002). The Arabidopsis model is likely to be very different from legumes in terms of responses to stress in relation to grain filling, nitrogen utilization, fixation, and transport, root architecture, and interactions, all physiological processes that are fundamentally different in legumes. Hence, the usefulness of developing a legume model has become increasingly relevant in recent years. In legumes, the gene expression patterns following biotic stresses have been more extensively studied than those following abiotic stresses. With respect to abiotic stress, gene expression analyses have been mainly based on studies with cloned genes (Singh et al. 2004). Significant progress is being made at the genetic and genomic levels using the model legume *M. truncatula* through macro- and microarray analysis, reverse genetics, genome sequencing, and other high-throughput techniques (Thompson et al. 2005; Oldroyd 2005). The analysis of almost 200,000 ESTs of *M. truncatula*, isolated from many different libraries constructed from diverse stages and treatments, was facilitated by searchable databases such as MtDB2 (Lamblin et al. 2003) and the TIGR Gene Index (http://www.tigr.org). Recent reports have also shown that transcriptomic tools are a good option for legume breeding to abiotic and biotic stresses.

Proteomics

Proteins act directly on biochemical processes, and thus must be closer to the phenotype, compared to DNA-based markers. The studies on proteins expressed by the genome of a cell, tissue, or organism at a specific time (proteome) is necessary to understand the biological function of a cell or an organism. Although research on plant responses to stress on the DNA or RNA level provided an important insight into stress tolerance, the proteomics approach is very important in evaluating stress responses since the mRNA levels may not always correlate with protein accumulation (Gygi et al. 1999). In addition, many proteins are modified by posttranslational modifications such as phosphorylation, glucosylation, and ubiquitinylation, which significantly influence protein functions. Proteomics, understood as protein biochemistry on an unprecedented and high-throughput scale, is becoming a promising and active approach in this postgenomic period. However, its application to plants is rather limited compared to other biological systems (Jorrin et al. 2006), although

good technical progress has been achieved in the separation of proteins and their identification by mass spectrometry. Studies have evaluated changes in protein levels in plant tissues in response to stresses (Canovas et al. 2004; Kim et al. 2003). However, these studies have mainly focused on nonlegume species such as Arabidopsis and rice (Canovas et al. 2004) and some legumes recently (Jorrin et al. 2006). As a result, only a handful of studies have been carried out in legumes, although in the next few years there should be a significant increase in the number of legume species. So far, pea has been more intensively studied, with the analysis of induced protein expression in roots in response to salt (Kav et al. 2004) and to cadmium stress (Repetto et al. 2003). Recently, *M. truncatula* has been the subject of several proteomic studies that represent the most extensive proteomic description to date and provide a reference map for future comparative proteomics and functional genomics studies of biotic and abiotic stress responses on legumes (Lei et al. 2005).

Metabolomics

Metabolomics provide the most direct tools for the quantitative measurement of the metabolism in an organism. Transcriptomic and proteomic data are important in deciphering a complex biological process, but they are still insufficient since most biological processes are ultimately mediated by cell metabolites. Thus, the complete understanding of both gene function and molecular events controlling complex plant processes requires analysis of transcriptome, the proteome, and the metabolome in an integrative manner (Dixon 2001). Metabolite profiling and metabolic fingerprinting are two different approaches in metabolomics that can be used for a large range of applications, including phenotyping of genetically modified plants, substantial equivalence testing, determining gene function, and monitoring responses to biotic and abiotic stresses. Recently, a promising platforms for metabolomics has been developed with the combination of two-dimensional gas chromatography and fast acquisition rate mass spectrometry (Scherling et al. 2010). Due to their specific technology, both technologies provide a complementary view of the metabolome such as amino acids, sugars, organic acids, free fatty acids, etc. However, most of the metabolomics platforms still need further method validation and quality checks. This is an essential requirement to guarantee meaningful biological applications. Furthermore, databases, experimental standards and data exchangeability between labs is an urgent issue for further developments in metabolomics (Weckwerth 2011a, b; Sansone et al. 2007).

In legumes, the metabolomic approach has been used in *M. truncatula* to determine the responses to various stimuli (Bell et al. 2001). Although, large-scale comprehensive metabolomic studies are difficult, a number of targeted analyses have been performed to assess the involvement of subsets of metabolites in various stresses. Although the preliminary results from combining metabolic approaches with transgenics indicates the potential of increasing intrinsic stress resistance levels in legume crops and strengthens the potential role of "omic" research in crop improvement (He and Dixon 2000; Wu and Van Etten 2004), it must be emphasized

that most metabolic pathways are interconnected in highly complex networks. Thus, modulating one metabolic pathway may have negative impacts on another, leading to concomitant deleterious traits in the modified crop. Large-scale metabolic analyses are therefore necessary to observe the metabolic networks important for plant growth and development under a range of environmental conditions.

Bioinformatics Platform

"Omic" era in the twenty-first century gives us opportunities to understand the legume genome at sequence-structural-functional levels. The rapid development of various genomic tools and techniques including large scale analysis of genome organization, gene expression, protein-protein interaction etc. are generating enormous amount of data which need to be analyzed and interpreted to develop a biologically meaningful concepts. The genome sequencing projects on different legumes generated the wealth of sequence data. These data need to be properly analysed to enable prediction of the potential functional elements, genes and transcription factors. Bioinformatics tools and databases help us in the analysis as well as understanding of the various features of the sequenced genome (Kushwaha et al. 2011; Dutt et al. 2010; Kumari et al. 2010). The availability of different biological databases related to legumes provides valuable information resource for research and analysis. Illustrated Legume Genetic Resources Database (www.gene.affrc.go.jp), LegumeTFDB (www.legumetfdb.psc.riken.jp) Bioinformatics resources for legume researchers (www.legumes.org), Chickpea Transcriptome Database (http://59.163.192.90:8080/ctdb/) are some of the important bioinformatics platforms providing important resources for legumes. These experimental datasets give us opportunities to understand the functional and biological roles of unknown genes/proteins from different legumes. Most of the assembler tools and packages were also developed e.g., short oligonucleotide analysis package and *de novo* assembly tools were developed by Beijing Genomics Institute (BGI). Several bioinformatics tools are available for annotation, genome sequence alignment, *de novo* assembly, sequence alignments and RNA sequence analysis. The basic level of genome annotation can be done using Basic Local Alignment Search Tool to find out similarities and differences. However, nowadays more and more additional information is added to the annotation platform. Bioinformatics tools developed for many non-legume species provides useful platform for legumes also.

Applications in Crop Improvement

All these technological platforms described above enable the genome-wide molecular analysis of different genotypes. This integrated high throughput analysis of metabolites, proteins and transcripts allow the definition of biochemical

phenotypes and their relationship to the corresponding genotype (Weckwerth 2008). The integration of metabolite and protein profiling has already been demonstrated to significantly improve pattern recognition and the selection and interpretation of multiple physiological and biomarkers for plant systems and different plant genotypes under different environmental conditions such as day-night rhythms or cold stress (Morgenthal et al. 2005; Wienkoop et al. 2008). Integration of metabolite and transcript data was also demonstrated to reveal the relationship between mRNA expression and dynamics of secondary metabolism (Tohge et al. 2005). The exploitation of these technologies in QTL-based marker-assisted breeding approaches (Fernie and Schauer 2009; Collard and Mackill 2008) is an obvious development. Most of the studies are focused on DNA markers. In recent studies the successful application of these technologies was also demonstrated for proteomics and metabolomics. Such efforts need to be accelerated for legumes. Marker-assisted selection could accelerate this process for the identification of useful traits in the early years of the selection process. It is anticipated that new technologies such as genomics, proteomics and metabolomics will yield such marker systems, however, these technologies have hardly reached the stage of application for breeding. A multitude of diagnostic marker assays will therefore be required for marker-assisted selection (Gebhardt et al. 2006). However, the robustness of these markers must be analyzed with higher statistical power from a higher number of samples. Both data sets—the metabolomics data and the proteomics data—show a good cultivar discrimination, however, the sample pattern can be interpreted differently depending on the characteristics of the different cultivars. Thus, the metabolite data carry different information to the protein data. Integration of these data leads consequently to optimized pattern recognition processes and improved interpretation of the molecular data with respect to the molecular phenotype which was indeed observed in several previous studies (Morgenthal et al. 2005; Wienkoop et al. 2008).

Prospects

Legumes are important for nutritional security and ecological sustainability. Exploitation of natural variation, population dynamics and a better understanding of the genotype-phenotype relationship is very crucial for improving productivity and stress adaptation of legumes. The development of genomic research during last decade may lead to a refinement of classical and molecular breeding approaches for legume improvement. A quite progress has been made for genome-scale investigation of some legumes to understand adaptation mechanisms and to provide fundamental knowledge for genetic variation, also for trait selection and genome/marker-assisted breeding approaches. Genomic resources provide the breeders a platform for rapid realization of resistance breeding objectives. A crucial pre-requisite for the deployment of markers to support stress tolerance or resistance breeding is the development of a genetic map, followed by identification of gene-based or gene linked markers to be used in marker assisted selection (MAS). Basic requirement of

availability of genomic resources for successful application of molecular markers in most of the legumes are now in place. Sequencing efforts have already made their strides in complete genome sequencing of few legumes like *M. truncatula*, *L. japonicus* and *Glycine max*, and draft genome sequencing of *Cicer arietinum* and *Cajanus cajan* genomes and the genome sequencing projects of some other legumes. This will generate large scale SNPs, SSRs and intron length polymorphic markers, which can help to saturate the linkage maps. Current genetic linkage maps of most of the legumes display an inadequate level of marker density. To improve the utility of such maps, it will be required to further saturate the map with additional markers.

Large-scale analysis by using different omics technologies are providing extensive data sets that will help identify potential candidate genes for enhanced productivity and stress adaptations in important legume crops. Identification of these candidate genes may allow their direct application in crop improvement through marker assisted breeding. However, in most cases, the roles of these candidate genes remain unknown and it will be important to carry out functional studies as a preliminary step toward their use in genetic improvement. The traditional pursuit of a gene starting with a phenotype (forward genetics) has paved the way for the opposite situation where the gene sequences are known but not their functions. The challenge is to decipher the function of thousands of genes identified by genome projects where reverse genetics methodologies will be the key tools. The ability to knockout genes or suppress their expression are powerful tools to determine the function of a gene. This can be done by antisense RNA suppression, targeted gene replacement, insertional mutagenesis, gene silencing through RNAi, and targeted induced local lesion in genome (TILLING) approaches.

Successful application of omics to legume improvement requires knowledge of stress response at molecular level, which includes gene expression to protein or metabolite and its phenotypic effects. Availability of genome sequence of legumes has a potential to facilitate positional cloning and other approaches and their applications for legumes research. A genome-wide expression profiling with next-generation sequencing approaches could circumvent the various problems in studying the legume genome. Compared to analysis of the transcriptome, analysis of the plant proteome and metabolome in response to abiotic and biotic stresses is still limited to *M. truncatula* and protein reference maps of soybean to stress responses are now available. More recently, few proteomics studies are available on chickpea and groundnut and they have to be extensively carried out for other legumes. Moreover, the recent progress in the mass-scale profiling of the genome, transcriptome, proteome, and metabolome offers the possibility of investigating the concerted response of thousands of genes to biotic and abiotic stresses. The mapping of abiotic stress QTL in legume is still at an early stage and gene pyramiding has not been applied yet. Nevertheless, with the establishment of the model legumes, *M. truncatula* and *L. japonicus*, there is now applicable information on legumes. Among the grain legumes, soybean has been more intensively studied, and the availability of more numbers of ESTs and genome sequences will facilitate mapping of major QTL in other legumes. Rapid progress in legume improvement will be possible with identification of candidate genes for desired traits in legumes. It is now possible

to target almost all legume crops with a variety of omics approaches for genetic improvement. All these efforts will lead to enhanced crop productivity of legume and ensure progress towards attaining nutritional security and ecological sustainability.

References

Asamizu E, Nakamura Y, Sato S, Tabata S (2004) Characteristics of the *Lotus japonicus* gene repertoire deduced from large-scale expressed sequence tag (EST) analysis. Plant Mol Biol 54:405–414

Bell CJ, Dixon RA, Farmer AD, Flores R, Inman J, Gonzales RA, Harrison MJ, Paiva NL, Scott AD, Weller JW, May GD (2001) The *Medicago* Genome Initiative: a model legume database. Nucleic Acids Res 29:114–117

Brautigam A, Gowik U (2010) What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. Plant Biol (Stuttg) 12:831–841

Canovas F, Dumas-Gaudot E, Recorbet G, Jorrin J, Mock HP, Rossignol M (2004) Plant proteome analysis. Proteomics 4:285–298

Castellana NE, Payne SH, Shen ZX, Stanke M et al (2008) Discovery and revision of Arabidopsis genes by proteogenomics. Proc Natl Acad Sci U S A 105:21034–21038

Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc Lond B Biol Sci 363:557–572

Cook DR (1999) *Medicago truncatula*: a model in the making! Commentary. Curr Opin Plant Biol 2:301–304

Dita MA, Rispail N, Prats E, Rubiales D, Singh KB (2006) Biotechnology approaches to overcome biotic and abiotic stress constraints in legumes. Euphytica 147:1–24. doi:10.1007/s10681-006-6156-9

Dixon RA (2001) Natural products and plant disease resistance. Nature 411:843–847

Doyle JJ (2001) Leguminosae. In: Brenner S, Miller JH (eds) Encyclopedia of Genetics. Academic, San Diego, pp 1081–1085

Dutt S, Singh VK, Marla SS, Kumar A (2010) *In silico* analysis of sequential, structural and functional diversity of wheat cystatins and its implication in plant defense. Genomics Proteomics Bioinformatics 8(1):42–56, PMID: 20451161

Fernie AR, Schauer N (2009) Metabolomics-assisted breeding: a viable option for crop improvement? Trends Genet 25(1):39–48

Gebhardt C, Bellin D, Henselewski H, Lehmann W et al (2006) Marker-assisted combination of major genes for pathogen resistance in potato. Theor Appl Genet 112:1458–1464

Goff SA, Ricke D, Lan TH, Presting G, Wang RL, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H et al (2002) A draft sequence of the rice genome (*Oryza sativa* L ssp *japonica*). Science 296:92–100

Gygi SP, Rochon Y, Franza BR, Aebersold R (1999) Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 19:1720–1730

Handberg K, Stougaard J (1992) *Lotus japonicus*: an autogamous, diploid legume species for classical and molecular genetics. Plant J 2:487–496

He XZ, Dixon RA (2000) Genetic manipulation of isoflavone 7-Omethyltransferase enhances biosynthesis of 40-O-methylated isoflavonoid phytoalexins and disease resistance in alfalfa. Plant Cell 12:1689–1702

Ideker T, Galitski T, Hood L (2001) A new approach to decoding life: systems biology. Annu Rev Genomics Hum Genet 2:343–372

Ishitani M, Rao I, Wenzl P, Beebe S, Tohme J (2004) Integration of genomics approach with traditional breeding towards improving abiotic stress adaptation: drought and aluminum toxicity as case studies. Field Crop Res 90:35–45

Jorrin JV, Rubiales D, Dumas-Gaudot E, Recorbet G, Maldonado A, Castillejo MA, Curto M (2006) Proteomics: a promising approaches to study biotic interaction in legumes: a review. Euphytica 147:37–47

Kalo P, Seres A, Taylor SA, Jakab J, Kevei Z, Kereszt A, Endre G, Ellis THN, Kiss GB (2004) Comparative mapping between *Medicago sativa* and *Pisum sativum*. Mol Genet Genomics 272L:235–246

Kato T, Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S (2003) Structural analysis of a Lotus japonicus genome V sequence features and mapping of sixty-four TAC clones which cover the 64 Mb regions of the genome. DNA Res 10:277–285

Kav NNV, Srivastava S, Goonewardene L, Blade SF (2004) Proteome-level changes in the roots of *Pisum sativum* in response to salinity. Ann Appl Biol 145:217–230

Kim ST, Cho KS, Yu S, Kim SG, Hong JC, Han CD, Bae DW, Nam MH, Kang KY (2003) Proteomic analysis of differentially expressed proteins induced by rice blast fungus and elicitor in suspension cultured rice cells. Proteomics 3:2368–2378

Kulikova O, Gualtieri G, Geurts R, Kim DJ, Cook D, Huguet T, de Jong JH, Franszn PF, Bisseling T (2001) Integration of the FISH pachytene and genetic maps of *Medicago truncatula*. Plant J 27:49–58

Kumari A, Singh VK, Fitter J, Polen T, Kayastha AM (2010) Alpha-amylase from germinating soybean (*Glycine max*) seeds-purification, characterization and sequential similarity of conserved and catalytic amino acid residues. Phytochemistry 71(14–15):1657–1666, PMID: 20655076

Kushwaha H, Gupta S, Singh VK, Rastogi S, Yadav D (2011) Genome wide identification of D of transcription factor gene family in sorghum and its comparative phylogenetic analysis with rice and *Arabidopsis*. Mol Biol Rep 38(8):5037–5053, PMID: 21161392

Lamblin AFJ, Crow JA, Johnson JE et al (2003) MtDB: a database for personalized data mining of the model legume *Medicago truncatula* transcriptome. Nucleic Acids Res 31:196–201

Lei Z, Elmer AM, Watson BS, Dixon RA, Mendes PJ, Sumner LW (2005) A two-dimensional electrophoresis proteomic reference map and systematic identification of 1367 protein from a cell suspension culture of the model legume *Medicago truncatula*. Mol Cell Proteomics 4:1812–1825

Lin XY, Kaul SS, Rounsley S, Shea TP, Benito ML, Town CD, Fujii CY, Mason T, Bowman CL, Barnstead M et al (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature 402:761

Mayer K, Schuller C, Wambutt R, Murphy G, Volckaert G, Pohl T, Dusterhoft A, Stiekema W, Entian KD, Terryn N et al (1999) Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. Nature 402:769

Moe KT, Chung JW, Cho Y, Moon JK, Ku JH, Jung JK, Lee J, Park YJ (2011) Sequence information on simple sequence repeats and single nucleotide polymorphisms through transcriptome analysis of mungbean. J Integr Plant Biol 53:67–73

Morgenthal K, Wienkoop S, Scholz M, Selbig J, Weckwerth W (2005) Correlative GC–TOF–MS based metabolite profiling and LC–MS based protein profiling reveal time-related systemic regulation of metabolite–protein networks and improve pattern recognition for multiple biomarker selection. Metabolomics 1:109–121

Ohtsu K, Smith MB, Emrich SJ, Borsuk LA, Zhou R, Chen T, Zhang X, Timmermans MCP, Beck J, Buckner B, Buckner DJ, Nettleton D, Scanlon MJ, Schnable PS (2007) Global gene expression analysis of the shoot apical meristem of maize (*Zea mays* L). Plant J 52:391–404

Oldroyd G (2005) Sequencing the model legume *Medicago truncatula*. Grain Legumes 41:23

Pedrosa A, Sandal N, Stougaard J, Schweizer D, Bachmair A (2002) Chromosomal map of the model legume *Lotus japonicus*. Genetics 161:1661–1672

Repetto O, Bestel-Corre G, Dumas-Gaudot E, Berta G, Gianinazzi-Pearson V, Gianinazzi S (2003) Targeted proteomics to identify cadmium-induced protein modifications in *Glomus mosseae* inoculated pea roots. New Phytol 157:555–567

Sansone SA, Fan T, Goodacre R, Griffin JL et al (2007) The metabolomics standards initiative. Nat Biotechnol 25:846–848

Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K et al (2008) Genome structure of the legume, *Lotus japonicus*. DNA Res 15:227–239

Scherling C, Roscher C, Giavalisco P, Schulze ED, Weckwerth W (2010) Metabolomics unravel contrasting effects of biodiversity on the performance of individual plant species. PLoS One 5:e12569

Schmutz J, Cannon S, Schlueter J, Ma J, Mitros T, Nelson W, Hyten D, Song Q, Thelen J, Cheng J, Xu D, Hellsten U, May G, Yu Y, Sakurai T, Umezawa T, Bhattacharyya M, Sandhu D et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Akurai T et al (2002) Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold, and high salinity stresses using a full-length cDNA microarray. Plant J 31:279–292

Shultz L, Kazi S, Bashir R, Afzal JA, Lightfoot DA (2007) The development of BAC-end sequence-based microsatellite markers and placement in the physical and genetic maps of soybean. Theor Appl Genet 114:1081–1090

Singh BN, Mishra RN, Agarwal PK, Goswami M, Nair S, Sopory SK, Reddy MK (2004) A pea chloroplast translation elongation factor that is regulated by abiotic factors. Biochem Biophys Res Commun 320:523–530

Singh NK, Gupta DK, Jayaswal PK, Mahato AK, Dutta S, Singh S, Bhutani S et al (2012) The first draft of the pigeonpea genome sequence. J Plant Biochem Biotechnol 21(1):98–112

Somerville C, Dangl J (2000) Genomics. Plant biology in 2010. Science 290:2077–2078

Stracke S, Sato S, Sandal N, Koyama M, Kaneko T, Tabata S, Parniske M (2004) Exploitation of colinear relationships between the genomes of *Lotus japonicus*, *Pisum sativum* and *Arabidopsis thaliana*, for positional cloning of a legume symbiosis gene. Theor Appl Genet 108:442–449

Thompson R, Ratet P, Kuster H (2005) Identification of gene functions by applying TILLING and insertional mutagenesis strategies on microarray-based expression data. Grain Legumes 41:20–22

Thoquet P, Gherardi M, Journet EP, Kereszt A, Ane JM, Prosperi JM, Huguet T (2002) The molecular genetic linkage map of the model legume *Medicago truncatula*: an essential tool for comparative legume genomics and the isolation of agronomically important genes. BMC Plant Biol 2:1

Tohge T, Nishiyama Y, Hirai MY, Yano M et al (2005) Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. Plant J 42:218–235

Vance CP, Graham PH, Allan DL (2000) Biological nitrogen fixation, phosphorus: a critical future need. In: Pedrosa FO, Hungria M, Yates MG, Newton WE (eds) Nitrogen fixation: from molecules to crop productivity. Kluwer, Dordrecht, The Netherlands, pp 506–514

Varshney RK, Glaszmann JC, Leung H, Ribaut JM (2010) More genomic resources for less-studied crops. Trends Biotechnol 28:452–460

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MTA, Azam S et al (2011) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotechnol 30:83–89

Wang L, Li P, Brutnell TP (2010) Exploring plant transcriptomes using ultra high-throughput sequencing. Brief Funct Genomics 9:118–128

Weckwerth W (2003) Metabolomics in systems biology. Annu Rev Plant Biol 54:69–89

Weckwerth W (2008) Integration of metabolomics and proteomics in molecular plant physiology—coping with the complexity by data-dimensionality reduction. Physiol Plant 132:176–189

Weckwerth W (2011a) Green systems biology—from single genomes, proteomes and metabolomes to ecosystems research and biotechnology. J Proteomics 75:285–305

Weckwerth W (2011b) Unpredictability of metabolism—the key role of metabolomics science in combination with next-generation genome sequencing. Anal Bioanal Chem 400:1967–1978

Wienkoop S, Morgenthal K, Wolschin F, Scholz M et al (2008) Integration of metabolomic and proteomic phenotypes: analysis of data covariance dissects starch and RFO metabolism from low and high temperature compensation response in Arabidopsis thaliana. Mol Cell Proteomics 7:1725–1736

Wienkoop S, Weiss J, May P, Kempa S et al (2010) Targeted proteomics for Chlamydomonas reinhardtii combined with rapid subcellular protein fractionation, metabolomics and metabolic flux analyses. Mol Biosyst 6:1018–1031

Wu QD, Van Etten HD (2004) Introduction of plant and fungal genes into pea (*Pisum sativum* L) hairy roots reduces their ability to produce pisatin and affects their response to a fungal pathogen. Mol Plant Microbe Interact 17:798–804

Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KF, Gouzy J, Schoof H, Van de Peer Y, Proost S, Cook DR, Meyers BC, Spannagl M, Cheung F, De Mita S, Krishnakumar V et al (2011) The *Medicago* genome provides insight into the evolution of rhizobial symbioses. Nature 480:520–524

Yu J, Hu SN, Wang J, Wong GKS, Li SG, Liu B, Deng YJ, Dai L, Zhou Y, Zhang XQ et al (2002) A draft sequence of the rice genome (*Oryza sativa* L ssp *indica*). Science 296:79–92

Chapter 2
Advances in Functional Genomics in Legumes

Marc Libault and Rebecca Dickstein

Abstract Functional genomics encompasses RNA transcription, protein expression and metabolomics as well as forward and reverse genetics. In recent years several resources like transcriptomes, proteomes, metabolomes, regulatory elements and mutant libraries using TILLING methods have been developed for legumes. These provide powerful molecular resources to identify the legume genes playing essential roles in plant resistance to biotic and abiotic stresses, regulating protein and oil accumulation in seeds and controlling beneficial plant-microbe symbiosis. Functional genomic studies on model legumes as well as on legumes of economic interest have already provided valuable information for enhancing legume productivity and holds promise for the future.

Keywords Legumes • Functional genomics • Genome • Epigenome • Transcriptome • Proteome • Metabolome • Abiotic stresses • Biotic stresses

Introduction

Functional genomics contributes to molecular breeding by identifying the expressed genes, proteins and metabolites associated with specific traits. It also identifies genes associated with water use efficiency (Kang et al. 2011) including those associated with stomatal opening and closure, nitrogen use efficiency, genes that respond to high temperature stress, those associated with flowering and seed set, those that are activated in response to pathogen attack and multiple other traits that are critical

M. Libault, Ph.D. (✉)
Department of Microbiology and Plant Biology, University of Oklahoma,
770 Van Vleet Oval, Room 32, Norman, OK 73019, USA
e-mail: libaultm@ou.edu

R. Dickstein, Ph.D.
Department of Biological Sciences, University of North Texas, Denton, TX, USA

S. Gupta et al. (eds.), *Legumes in the Omic Era*, DOI 10.1007/978-1-4614-8370-0_2,

for crop improvement. In legumes, functional genomics is helping to identify the genes regulating legume development, yield, their resistance to biotic and abiotic stresses, and other factors affecting their economic value.

Functional genomics in legumes had its start in the late 1990s with the first publication of expressed sequence tags (ESTs) from root hairs and nodules in the model legumes, *Medicago truncatula* and *Lotus japonicus* (Covitz et al. 1998; Szczyglowski et al. 1997), followed by EST analyses in several tissues (Endo et al. 2000; Asamizu et al. 2000; Gyorgyey et al. 2000). Other EST analyses soon followed in *Medicago*, *Lotus*, *Glycine max* and other legumes, with tissues collected from various organs and at different developmental times in model legumes as well as crop legumes (Fedorova et al. 2002; Journet et al. 2002; Shoemaker et al. 2002; Sawbridge et al. 2003; Schroeyers et al. 2004). These ESTs were organized into tentative consensus (TC) sequences and were originally housed at TIGR (Lee et al. 2005); now the TCs are available at DFCI (http://compbio.dfci.harvard.edu/tgi).

Recently, functional genomic studies have been accelerated with the emergence of high-throughput technologies. These technologies have been applied to the characterization of physiological and molecular changes occurring in plants responding to environmental stresses, as well as between organs, tissues or even single cell types. It is assumed that the relative abundance of transcripts, proteins or metabolites would provide an important indication of their role in plant development or in the plant response to a stress.

High-throughput DNA sequencing technologies have been used to develop genetic tools and resources used daily by functional genomicists such as the generation of drafts of legumes genome sequences (e.g. (Schmutz et al. 2010; Young et al. 2011)), the identification of genetic markers such as SNPs (Single Nucleotide Polymorphisms (Hyten et al. 2010; Cortes et al. 2011; Han et al. 2011; Muchero et al. 2011; Shu et al. 2011; Varala et al. 2011; Xu et al. 2011)), the establishment of the epigenome (i.e., genomic DNA methylome profiles and the mapping of histone post-translational chemical modification in the genomic DNA; (Schmitz and Zhang 2011)) and the deep characterization of legume transcriptomes (Benedito et al. 2008; Hogslund et al. 2009; He et al. 2009; Libault et al. 2010a; Severin et al. 2010). Proteomic and metabolomics capabilities have also increased in terms of sensitivity to now allow researchers the ability to identify thousands of proteins and metabolites from smaller and smaller plant tissue sample sizes (e.g. (Watson et al. 2003; Farag et al. 2008; Brechenmacher et al. 2009)).

Development of Resources for Functional Genomics in Legumes

Transcriptomic Resources

Genes that are differentially expressed genes in response to a stress or across organs, tissues or cell types are candidates to have a role in the adaptation of the plant to the stress, in organ development or in functionality. Hence, the establishment of the

transcriptional patterns of genes is a valuable starting point to identify genes controlling a biological process. The quantification of the expression level of an organisms' set of genes, which is reflected by their mRNA abundance, leads to the establishment of the transcriptome of single cell types, tissues and organs.

Arrays have been extensively developed and used during the past years to study the transcriptome of model legumes (Kouchi et al. 2004; Kuster et al. 2004; Suganuma et al. 2004; Vodkin et al. 2004; Lohar et al. 2006; Jones et al. 2008; You et al. 2011; Takanashi et al. 2012; Zahaf et al. 2012). These arrays, which represent a collection of transcript-specific oligonucleotides, allow the discrimination and the quantification of the abundance of each transcript in an organism. In addition to being useful for same species comparisons, the arrays may be used across species, although caution must be taken. For example, taking advantage of the close evolutionary relationship between *G. max, Vigna unguiculata* and *Phaseolus vulgaris*, the soybean arrays were also used with success to quantify the expression of *P. vulgaris* genes, a non-model legume plant (Das et al. 2010; Yang et al. 2010). In addition to these "oligonucleotide" arrays, the Affymetrix Company has developed and commercialized arrays to characterize the expression pattern of *G. max* (Valdes-Lopez et al. 2011) and *M. truncatula* genes (Mitra et al. 2004a). *M. truncatula* arrays have been used to profile alfalfa genes (Kang et al. 2011). An Affymetrix array was used to generate the *M. truncatula* Gene Expression Atlas (MtGEA) (Benedito et al. 2008). This atlas groups a large number of transcriptomic analyses (Benedito et al. 2008; Naoumkina et al. 2007, 2008; Imin et al. 2008; Holmes et al. 2008; Ruffel et al. 2008; Uppalapati et al. 2009; Gomez et al. 2009; Pang et al. 2008) and is hosted by a webserver, allowing the analysis and manipulation of the transcriptomic data sets (http://mtgea.noble.org/v2/; (He et al. 2009)). Overall, the quality of the arrays and their coverage is highly dependent on the quality of the cDNA libraries used to design the oligonucleotides on the array. For example, due to the complexity of the soybean genome; i.e., its recent duplication 13 Mya, as well as the use of incomplete cDNA libraries, the first generation of the soybean Affymetrix array did not provide an optimal set of oligonucleotides, leading to a limited coverage of the soybean transcriptome as well as a lack of specificity of some of the oligonucleotides for specific transcripts (Libault et al. 2010b).

High-throughput sequencing recently became a reference technology to characterize legume transcriptomes. Various platforms exist, allowing the generation of different number of reads; i.e., sequencing products of different lengths (Fig. 2.1). Hence, biologists are now frequently using high-throughput sequencing technology to characterize legume transcriptomes. For example, the use of this technology led to the establishment of the soybean transcriptome atlases (Libault et al. 2010a; Severin et al. 2010). Coupled with the development of bioinformatics tools, these transcriptomic resources can be easily accessed from two different bioinformatics platforms: the Soyseq platform hosted on Soybase (http://soybase.org/soyseq/) and Soykb (http://soykb.org/; (Joshi et al. 2012)). The drop of the this technology's cost as well as its higher sensitivity and accuracy in measuring transcript abundance now allow scientists to use it to characterize gene expression patterns in legumes. In model legumes, use of high-throughput sequencing enables transcript abundance measurements of genes missing from microarray platforms. Also, since this

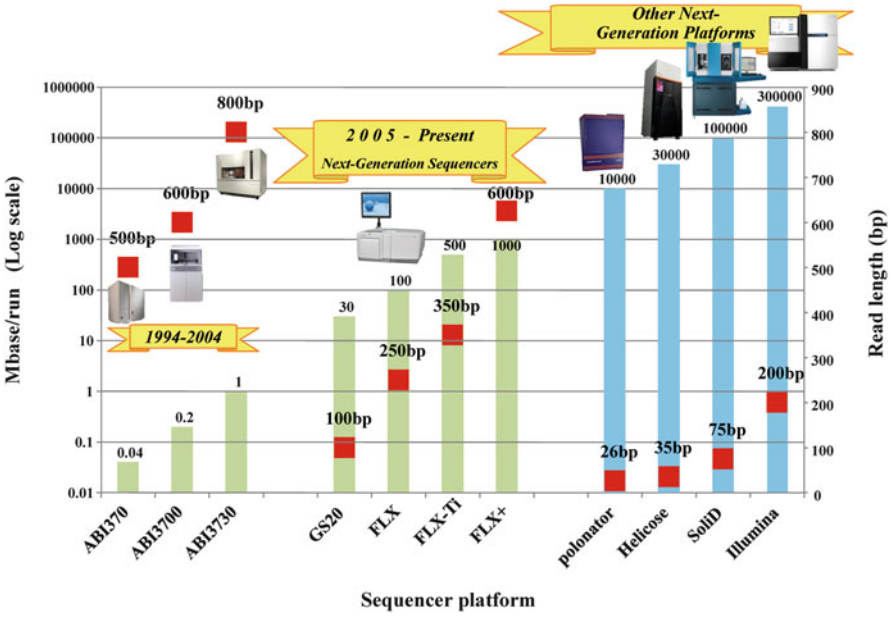


Fig. 2.1 A brief history of long read automated DNA sequencing instruments (Courtesy of Dr. Fares Najar from the University of Oklahoma)

technology is based on *de novo* sequencing of DNA fragments, another advantage is that it does not rely on the need for genome sequences to map the reads. Hence with the support of bioinformatics tools, which are required to assemble the sequences and create contings, transcriptomic data sets can also been generated in any legume (e.g. *Acacia auriculiformis* and *Acacia mangium* (Wang et al. 2011); *Maruca vitrata* (Margam et al. 2011); *P. vulgaris* (Kalavacharla et al. 2011); *Cicer arietinum* (Garg et al. 2011; Hiremath et al. 2011)).

Efforts have been made to create large scale libraries of oligonucleotide primers, allowing the quantification of *G. max* and *M. truncatula* transcription factor gene expression using quantitative RT-PCR reaction (qRT-PCR) (Libault et al. 2009; Kakar et al. 2008; Verdier et al. 2008). qRT-PCR, which is the most sensitive technology to accurately quantify gene expression, is a especially well-suited to quantify transcription factor gene expression. Transcription factors have overall low expression of these genes, making them less amenable to quantitation based on microarrays or other high-throughput strategies. Additionally, redundancy of some of their nucleotidic sequences requires the design of specific primer sets.

DNA microarray hybridization and high-throughput sequencing technologies have been used to characterize legume transcriptomes, resulting in the development of gene expression atlases. Similar to the development of Genevestigator which was developed for several systems including *Arabidopsis thaliana* (Hruz et al. 2008), a transcriptional platform developed for the model plant *A. thaliana*, the

development of bioinformatics tools allowing the individual investigator to query specific genes' or groups of genes' expression was a major milestone for the legume functional genomics community; i.e., the *M. truncatula* Gene Expression Atlas (Benedito et al. 2008; He et al. 2009) and the Soybean Transcriptome atlases (Libault et al. 2010a; Severin et al. 2010; Joshi et al. 2012); http://soybase.org/, http://soykb.org/.

Proteomes and Metabolomes

Legumes are an important source of protein used in animal and human nutrition especially in developing countries. Also, several metabolites, such as flavonoids, have been characterized for their beneficial impact on human health (Valachovicova et al. 2004). Hence, it is imperative to characterize the proteome and metabolome of legumes, with the goal of improving their nutritive value. As a consequence, metabolomic and proteomic approaches are of great interest to detect and quantify proteins and metabolites. Several studies characterized and integrated changes in legume metabolomes in response to biotic and abiotic stresses (Farag et al. 2008; Brechenmacher et al. 2010; Wu et al. 2008; Hernandez et al. 2009; Sanchez et al. 2010) as well as the protein composition in legume seeds (e.g., *M. truncatula* (Zhang et al. 2006; Gallardo et al. 2003; Gallardo et al. 2007; Repetto et al. 2008; Chatelain et al. 2012), *L. japonicus* (Dam et al. 2009; Nautrup-Pedersen et al. 2010), *G. max* (Hajduch et al. 2006; Agrawal et al. 2008; Krishnan et al. 2009), *P. vulgaris* (Marsolais et al. 2010) and *Pisum sativum* (Bourgeois et al. 2009)). The integration of the legume metabolomes and proteomes together with transcriptomes and genomes will improve our understanding of legume metabolomics pathways. As a consequence, bioinformatic tools have been created to analyze and integrate metabolomics and proteomic data sets, which is described in Table 2.1. These tools have been develop broadly to better understand plant biology [KEGG (http://www.genome.jp/kegg/); MetaCyc ((Caspi et al. 2010), http://metacyc.org/); Aracyc ((Mueller et al. 2003), http://www.arabidopsis.org/biocyc/)] but also especially developed to highlight the specificity of legume proteomes and metabolomes, e.g., The Soybean Proteome Database ((Sakata et al. 2009), http://proteome.dc.affrc.go.jp/Soybean/); Soykb ((Joshi et al. 2012), http://soykb.org/); MedicCyc ((Urbanczyk-Wochniak and Sumner 2007), http://pathway.gramene.org/gramene/mediccyc.shtml).

Regulatory Elements

Transcriptomic data along with genomic information, can also be used to identify the regulatory elements controlling gene expression. These regulatory elements are essential DNA promoter sequences recognized by the transcriptional regulatory proteins.

Table 2.1 Legume functional genomic bioinformatic tools

	Website	Database name
General information on legumes	http://www.nsrl.uiuc.edu/international.html	National Soybean Research Laboratory
	http://www.soybiotechcenter.org/	National Center for Soybean Biology
	http://www.comparative-legumes.org/	Legume Information System
Genomic resources	http://www.phytozome.net/	Phytozome
	http://www.medicagohapmap.org/	*Medicago truncatula* HapMap Project
	http://www.jcvi.org/cgi-bin/medicago/overview.cgi	*Medicago truncatula* Genome Project
	http://www.jcvi.org/cgi-bin/gb2/gbrowse/mtruncatula/	Medicago Genome Browser
	http://www.kazusa.or.jp/lotus/	Lotus genome database
Legume synteny	http://www.symapdb.org/	SyMAP Synteny Browser
	http://genomevolution.org/CoGe/	CoGe, The plant to Compare Genomes
	http://plantgrn.noble.org/LegumeIP/	LegumeIP
Transcriptomic resources	http://mtgea.noble.org/v2/	*Medicago truncatula* Gene Expression Atlas
	http://soybase.org/soyseq/	Soybase
	http://soykb.org/	SoyKB
Proteomic resources	http://soykb.org/	SoyKB
	http://proteome.dc.affrc.go.jp/Soybean/	Soybean Proteome Database
Metabolomic resources	http://soykb.org/	SoyKB
	http://soybase.org/soyseq/	Soybase
	http://plantcyc.org/	Plant Metabolic Network
	http://pathway.gramene.org/gramene/mediccyc.shtml	Medicago Metabolic Pathways
TILLING resources	http://revgenuk.jic.ac.uk/TILLING.htm	RevGen UK
	http://www.inra.fr/legumbase	Legume base
	http://www.soybeantilling.org	Soybean Mutation Database
Functional genomic	http://compbio.dfci.harvard.edu/tgi	Computational Biology and Functional Genomics Laboratory
	http://plantgrn.noble.org/LegumeIP/	LegumeIP
	http://bioinfo3.noble.org/medicago/index_MT3.html	Medicago genome portal at the Noble Foundation
	http://www.genome.jp/kegg/	Kyoto Encyclopedia of Genes and Genomes
Genetic and physical maps	http://soybase.org/soyseq/	Soybase

The characterization of these regulatory networks is difficult and only a few studies could be successful to unravel these elements. Among them, several studies characterized the regulatory elements of the *MtENOD11* (*Early NODulin 11*) gene previously characterized for its early induction of its expression during nodulation.

The *MtENOD11* Nod Factor-Responsive Element interacting with the ERF transcription factor (Andriankaja et al. 2007) and the Nodulation Responsive Elements 1 and 2 recognized by the NSP1 GRAS transcription factor were identified (Hirsch et al. 2009). This latter study also highlighted the interaction between NSP1 and NSP2 GRAS transcription factors as an essential component activating *MtENOD11* gene expression. The characterization of *cis*-acting regulatory elements can now been accessed using chromatin-immunoprecipitation methods (ChIP) coupled with high-throughput sequencing technology. This is called the ChIP-Seq method (Muino et al. 2011). Complementary to these approaches, bioinformatics tools have been developed to reconstruct gene regulatory networks (Wang et al. 2010).

It is now clear that the characterization of the interactions between transcription factors and their promoter binding sites is not sufficient to fully explain the regulation of eukaryotic gene expression. In fact, the regulation of gene expression is not only controlled by the interaction between *cis*-regulatory elements and transcription factors but also by reversible chemical modifications of the gDNA, e.g., methylation of cytosine residues and post-translational modification of histones, proteins associated to the gDNA (for review, (Schmitz and Zhang 2011)). These chemical modifications are named as epigenetic marks. The epigenome refers to the overall composition of these epigenetic marks. Until very recently, the impact of the epigenome controlling gene expression and regulating legume biological processes has been neglected. Taking advantage of the sensitivity of high-throughput sequencing technology as well as the development of bioinformatics tools (Sunkar and Jagadeeswaran 2008; Zhang et al. 2008, 2010; Zhou et al. 2008; Xuan et al. 2011), legume micro RNAs were identified (Wong et al. 2011; Subramanian et al. 2008; Szittya et al. 2008; Jagadeeswaran et al. 2009; Joshi et al. 2010; Zhe et al. 2013; Devers et al. 2011; Kulcheski et al. 2011; Li et al. 2010; Song et al. 2011; Wang et al. 2011; Chen et al. 2012a, 2012b; Zhou et al. 2012) and efforts to characterize their putative targeted genes and function have been initiated. In addition, several studies have identified small RNA from non-model legumes such as peanut (*Arachis hypogaea*, (Zhao et al. 2010)), common bean (*P. vulgaris*, (Valdes-Lopez et al. 2008; Arenas-Huertero et al. 2009)) and wild soja (*Glycine soja*, (Chen et al. 2009)). In soybean, Li et al. (2010) highlighted the role of miR482, miR1512, and miR1515 as key regulators of nodulation while the role of *M. truncatula* small RNAs in controlling nodulation (miRNA166 (Boualem et al. 2008); miRNA169 (Combier et al. 2006)), root development (miRNA166 (Boualem et al. 2008)), plant response to inorganic phosphate during mycorrhization (miRNA399 (Branscheid et al. 2010)) and to water deficit (miR398 and miR408 (Trindade et al. 2010)) have been demonstrated. While consistent effort has been made during the past 5 years to enumerate and characterize the function of small RNAs in legume biology, other epigenetic regulatory mechanisms have been poorly explored in legumes. The proteomic analysis of soybean histone post-translational modifications clearly highlighted specific chemical modification of soybean histones compared to *A. thaliana* histones, e.g., methylation of H3K79 (Wu et al. 2009). However, the mapping at the scale of legume genomes of histone post translational modifications and cytosine methylation is not currently established in soybean organs, tissues, cells.

Mutant Libraries

Reverse genetics in legumes, an important tool in functional genomics, requires mutagenized legume populations that can be screened for variation, deletion or insertions in the genes of interest. To fully unleash the benefits of the sequencing of the genome of the three legume models, mutant collections have been developed during the past years (Cook 1999; Penmetsa and Cook 2000; Catoira et al. 2000; Schauser et al. 1999). In the case of *M. truncatula* and *L. japonicus*, mutant collections were developed before the sequencing their genomes. A *Tnt1* retrotransposon insertional *M. truncatula* mutant collection is available at the Samuel Roberts Noble Foundation (Ardmore, OK, USA; (Tadege et al. 2008)) while an ethylmethanesulphonate (EMS) *L. japonicus* mutant collection was generated at the John Innes Center (UK) (Perry et al. 2009). In soybean, mutant libraries are under construction; i.e., the group of Wayne Parrott at the University of Georgia is leading the generation of transposon-based mutants using the tobacco retrotransposon *Tnt1*, the maize *Ac/Ds* transposon and the rice miniature inverted terminal repeat element (MITE) transposon *mPing* (Hancock et al. 2011).

In addition to these resources, TILLING (Targeting Induced Local Lesions IN Genomes; (Colbert et al. 2001)) is currently used to identify mutations within specific, targeted genes, leading to characterization of gene function. This technology is based on the use of the plant endonuclease CEL I which cleaves specifically heteroduplex mismatched sites (Kulinski et al. 2000) This technology allows the detection of mismatches between annealed wild-type and mutant DNA strands. Hence, mutagenesis, usually by the point mutation mutagen ethyl methane sulfonate (EMS), is a pre-requirement to TILLING experiments. The TILLING strategy has been applied to *M. truncatula*, *L. japonicus* and *G. max* mutant EMS populations. Developed resources for TILLING are now available at the John Innes Center (UK; http://revgenuk.jic.ac.uk/TILLING.htm; (Perry et al. 2003, 2009)), the Institut National de la Recherche Agronomique (France; http://www.inra.fr/legumbase; (Le Signor et al. 2009)) and Southern Illinois University (http://www.soybeantilling.org/; (Cooper et al. 2008)). Mutants identified via the TILLING strategy have been used successfully to characterize the function of legume genes. For example, taking advantage of TILLING technology, Dierking and Bilyeu (2009) identified mutations in soybean raffinose synthase genes, while Ariel et al. (2010) characterized the function of the HD-Zip I transcription factor HB1 in the emergence of *M. truncatula* lateral roots.

Fast neutron mutagenesis, which induces deletion mutations, has been applied with success to characterize *M. truncatula*, *L. japonicus* and *G. max* mutants showing defects in nodulation (Starker et al. 2006; Hoffmann et al. 2007; Bolon et al. 2011; Murray et al. 2011) as well as in seed development (Bolon et al. 2011). Fast neutron mutagenesis, when used together with comparative genomic hybridization (CGH) technology allow the mapping of the deletion(s) at the scale of the entire genome. The combination of fast neutron mutagenesis and CGH is powerful to identify candidate genes associated with phenotypes of interest. Fast neutron

mutagenesis has also been coupled with TILLING strategies leading to the development of a new reverse genetic strategy named Deletion TILLING (De-TILLING). De-TILLING has been successfully used to identify mutants in the *MtEFD* gene encoding an ethylene response transcription factor previously characterized for being up-regulated during *M. truncatula* nodulation (El Yahyaoui et al. 2004) and controlling *M. truncatula* nodulation (Vernie et al. 2008).

Use of Functional Genomics in Legume Improvement

Seed Composition

Using functional genomics, the genes could be indentified which were important for metabolic and regulatory networks leading to seed development. Investigations were made in soybean and *M. truncatula* embryo and seed development (Benedito et al. 2008; Joshi et al. 2012; Verdier et al. 2008; Le et al. 2007; Wang et al. 2012; Buitink et al. 2006). Functional genomics were also used in conjunction with legume mutant populations to identify important legume seed genes. Medicago populations have been mutagenized with the retrotransposons *Tnt1* (Tadege et al. 2005, 2008; Benlloch et al. 2006; Cheng et al. 2011) and *MERE1* (Rakocevic et al. 2009), by fast neutrons (FN) and X-rays causing deletions (Starker et al. 2006; Rogers et al. 2009; Sagan et al. 1995) and by agents that cause point mutations (Catoira et al. 2000; Le Signor et al. 2009; Porceddu et al. 2008). Soybean populations with point mutations have been constructed and screened by TILLING for mutations in genes associated with seed quality. Plants were found containing mutations in targeted genes and some of these had altered seed composition (Dierking and Bilyeu 2009). In a different study, a soybean fast neutron mutant population was screened with changes in seed protein and oil composition. Eight mutants were identified and subjected to comparative genome hybridization using a custom NimbleGen microarray designed to contain 696,139 gene probes, approximately one every 1.1 kb along the genome. Deleted regions were identified, revealing candidate genes associated with the seed quality phenotypes (Bolon et al. 2011). Functional genomics has been used to modify soybean to contain increased levels of the phytosterol sitosterol, desirable for oilseeds and human health (Neelakandan et al. 2012). A high-density oligonucleotide microarray was constructed and used to identify tissue-specific genes from peanut pods. These were found responsible for seed storage proteins as well as desiccation. Many transcripts highly induced in pods were previously unknown, now allowing for the possibility of assigning function to uncharacterized genes (Payton et al. 2009). Functional genomic studies in seeds may also impact other tissues in legumes. For example, in *M. truncatula*, qRT-PCR studies were used to identify transcription factors (TFs) activated during seed development (Verdier et al. 2008). Reverse genetics identified *M. truncatula Tnt1* mutants in one of these TFs, MtPAR, a MYB-type TF, that resulted in seeds

deficient in proanthocyanidin (PA) biosynthesis. *MtPAR* is normally expressed exclusively in the seed coat, where PA accumulates. Ectopic *MtPAR* expression in alfalfa (*M. sativa*) resulted in shoots that contained detectable levels of PA, opening the potential for *MtPAR* expression in forage legumes to contain PAs in foliage, which may reduce pasture bloat in ruminant animals (Verdier et al. 2012).

Resistance Against Pathogens

Legumes are susceptible to pathogen attack by a number of soil-borne pathogens. Among them are *Phymatotrichum* rot, sometimes called cotton or Texas root rot, attacking diverse species including alfalfa (Uppalapati et al. 2010). Extensive cytological and biochemical research aimed at understanding how the fungus penetrates and invades root tissue has recently been augmented with a functional genomics approach using the model legume *M. truncatula*, revealing that jasmonic acid, ethylene and flavonoid pathways are involved in disease development (Uppalapati et al. 2009). It has been determined that *M. truncatula* is susceptible to *Fusarium* wilt (Ramírez-Suero et al. 2010) and charcoal rot (Gaige et al. 2011) and future studies on these pathogens' mechanisms of attack in Medicago may include functional genomics approaches. In soybean, sudden death syndrome (SDS) is caused by *Fusarium* species and macroarray analysis as well as transcriptional profiling of susceptible and resistant soybean genotypes has been carried out. In the macroarray study, relatively large differences were noted in transcript abundance in susceptible and partially resistant soybean lines (Iqbal et al. 2005). In the latter profiling experiments, although many of the genes determined to increase in expression were similar in susceptible versus resistant lines, there were genotype-specific expression differences that were compared to known quantitative trait loci, narrowing candidate genes that control SDS-defense (Radwan et al. 2011).

Asian soybean rust (ASR), caused by *Phakopsora pachyrhizi*, an obligate, biotrophic, plant-pathogenic fungus, was recently introduced into the major soybean-growing countries of the Western Hemisphere. This event has generated considerable interest in the molecular interactions of *P. pachyrhizi* with its soybean host as well as with nonhost plants, some of which have been studied using functional genomics approaches (Choi et al. 2008; Schneider et al. 2011; Garcia et al. 2008; van de Mortel et al. 2007; Panthee et al. 2009; Soria-Guerra et al. 2010a, b; Tremblay et al. 2010). Medicago is being used as a model for pathogens that affect foliage. *M. truncatula* genotypes differentially susceptible to *Phoma medicaginis* have been studied using functional genomics, demonstrating that the octadecanoid and phenylpropanoid pathways are stimulated by this pathogen (Kamphuis et al. 2010). Earlier, it was shown by profiling flavonoid glycoconjugates that rapid increase of these molecules was correlated to the infection process in *M. truncatula* (Jasinski et al. 2009). *M. truncatula* is also being used as a model for alfalfa rust, incited by *Uromyces striatus*, and recently transcription factors (TFs) differentially expressed in resistant vs. susceptible *M. truncatula* genotypes

were profiled, with the resulting TFs mapped to chickpea using syntenic relationships (Madrid et al. 2010). Susceptible and resistant closely related lines of *M. truncatula* were assessed for TF expression following 6 and 12 h of bluegreen aphid (*Acyrthosiphon kondoi*) infestation and bluegreen aphid-induced expression of a subset was correlated with a single dominant gene conferring resistance to bluegreen aphids (Gao et al. 2010). Medicago is also susceptible to the soybean pathogen *P. pachyrhizi* and serves as a model for infection by *Puccinia emaculata*, the causative agent of switchgrass rust disease of switchgrass. Recently, a forward screen for *M. truncatula* mutants resistant to *P. pachyrhizi* and *P. emaculata* identified *IRG1/PALM1,* encoding a Cys(2)His(2) zinc finger transcription factor that had been previously identified as controlling leaf morphology, as also controlling leaf asymmetric epicuticular wax deposition, influencing fungal spore differentiation (Chen et al. 2010; Uppalapati et al. 2012). Transcription profiling in *irg1/palm1* mutants identified many mis-regulated genes predicted to function in wax/lipid biosynthesis (Uppalapati et al. 2012).

Resistance Against Nematodes

Root cyst and root knot nematodes are parasites that are among the world's most damaging crop pests. Soybean is an affected crop, and susceptible and resistant lines of soybean have been assessed for transcriptional responses both to cyst (Mazarei et al. 2011; Matsye et al. 2011; Klink et al. 2007a, 2010a, 2011; Ithal et al. 2007a, b; Puthoff et al. 2007) and root knot (Ibrahim et al. 2011) nematodes. Several elegant experiments coupled laser capture microdissection with microarray analysis in soybean resistant vs. susceptible lines or soybean near-isogenic lines differing at a major quantitative trait locus for resistance to cyst nematodes (Kandoth et al. 2011; Klink et al. 2007b); other studies have also coupled laser capture microdissection with transcriptome analyses (Klink et al. 2009a, 2010b). Proteomic and metabolic studies have also been done (Afzal et al. 2009). Together these studies identified many differentially expressed soybean genes and proteins, especially those associated with defense, in resistant lines compared to susceptible lines. Since genes encoded by soybean cyst nematode are on the soybean gene chip, some studies have identified differentially regulated cyst nematode genes (Klink et al. 2009b; Alkharouf et al. 2007). These identified genes are leading to new strategies for breeding and genetic engineering of new soybean lines and cultivars that are resistant to these damaging pathogens (Klink et al. 2009b; Alkharouf et al. 2007; Klink and Matthews 2009; Wise et al. 2007). Root-knot nematodes infect other legumes as well as soybean (Caillaud et al. 2008; Anderson et al. 2010; Wasson et al. 2009; Quesenberry et al. 2008; Poch et al. 2007; Stirling et al. 2006; Davis and Mitchum 2005; Lohar and Bird 2003). Some of these interactions have been studied using gene profiling using available gene chips from related legume species, for example, using the soybean microarray to study root knot nematode induced gene expression in cowpea (*Vigna unguiculata*) (Das et al. 2010).

Tolerance to Abiotic Stresses

Legumes deal with many abiotic stresses, some of which are being investigated by functional genomics approaches. Drought and salt stress are important abiotic stresses affecting legume production in tropics. In a drought and salt stress study in the model legume *L. japonicus*, metabolite profiling revealed gradual increases of many soluble small molecules, that were compared to similar stress in forage legume *Lotus* species. The results showed that only a few salt- and drought-responsive metabolites were common in all species examined (Sanchez et al. 2012). This is consistent with a similar transcriptomic and metabolomic study using model and cultivated *Lotus* species showing many genotype-specific transcriptional and metabolic changes but only a very small percentage of changes common to all species profiled (Sanchez et al. 2011). Both of these studies suggest caution in interpreting changes in gene expression and metabolites from only one species. An earlier study showed evidence for conserved and divergent metabolic responses to salinity in *L. japonicus* compared to two non-legume species (Sanchez et al. 2008a). A different earlier study that profiled at the ionomic, transcriptomic and metabolomic levels in *L. japonicus* during increasing salt stress suggested successive and global needs for gene expression and metabolic pathway reprogramming for maintenance of osmotic and ionic homeostasis (Sanchez et al. 2008b). TF profiling using quantitative RT-PCR of *M. truncatula* root apexes showed several TFs with large fold expression changes in salt stress conditions. A number of the TF genes also responded to other abiotic stresses, suggesting that they may participate in a general stress response, and thus are potential targets for mitigating stress (Gruber et al. 2009). The *M. truncatula* gene chip has been used to examine two alfalfa (*M. sativa*) varieties that differ in their tolerance/sensitivity to drought. The study, which also included a metabolomics analysis of the two varieties, showed differences in accumulation of osmolytes as well as differential regulation of TFs and other regulatory proteins between the two varieties, and thus it identified potential targets for improving drought tolerance in alfalfa (Kang et al. 2011). The results are consistent with an alfalfa leaf proteome study demonstrating that in drought conditions, plants invest a large quantity of resources into osmolytes to maintain turgor (Aranjuelo et al. 2011).

Phosphate (P) stress is a serious problem for legumes: P deficiency can have significant effects on N fixation, including nodule number and mass, nitrogenase activity and N content (Tang et al. 2001; Olivera et al. 2004; Sa and Israel 1991; Le Roux et al. 2006; Ribet and Drevon 1995; Vance et al. 2003). Functional genomic studies are identifying mechanisms by which legumes deal with P stress and identifying potential targets that may be useful in the future to help alleviate symptoms of P stress. Proteomic studies identified 44 P-starvation responsive proteins identified from soybean nodules and qRT-PCR verified that gene expression correlated with some of the P-regulated proteins (Chen et al. 2011). A number of important systems biology studies of P stress have been done in common bean, the most important human dietary legume, that is frequently cultivated in areas that lack sufficient P in the soil (Hernandez et al. 2007, 2009; Broughton et al. 2003; Graham et al. 2006). These studies, as well as others conducted in non-legumes have led to

the identification of genes and miRNAs important in regulating legume plants response to P stress as well as other abiotic stressors (Valdes-Lopez et al. 2008, 2010; Hernandez et al. 2009).

Root Symbioses

The legume root symbioses with soil rhizobia, producing nitrogen-fixing root nodules, and with mycorrhizal fungi, that help plants obtain nutrients from the soil, provide examples where functional genomics has led to many breakthroughs. Transcript profiling has identified many legume genes that are differentially regulated during the symbioses (Fedorova et al. 2002; Colebatch et al. 2002, 2004; Benedito et al. 2008, 2010; Hogslund et al. 2009; El Yahyaoui et al. 2004; Mitra et al. 2004b; Gomez et al. 2009; Liu et al. 2003, 2007; Guether et al. 2009; Hohnjec et al. 2005; Deguchi et al. 2007; Manthey et al. 2004; Asamizu et al. 2005; Brechenmacher et al. 2008; Mitra and Long 2004; Frenzel et al. 2005), including those encoding microRNAs (Subramanian et al. 2008; Joshi et al. 2010; Devers et al. 2011; Lelandais-Brière et al. 2009; Omrane et al. 2009; Simon et al. 2009; Zhai et al. 2011; Udvardi et al. 2007; Samac and Graham 2007). Transcript-based cloning has helped to identify genes essential to nitrogen-fixing root nodule and mycorrhizal root development, including a calcium calmodulin kinase *CCaMK*, the transcription factor genes *NSP2* and *ERN*, *GmFWL1*, associated with changes in chromatin structure, and *VAPRIN*, controlling infection by both rhizobia and AM fungi (Murray et al. 2011; Mitra et al. 2004b; Middleton et al. 2007; Libault et al. 2010c; Kalo et al. 2005). Soybean root hair cells, a single cell type, have been subjected to extensive functional analyses, including transcriptomics and metabolomics in the basal state and in response to its rhizobial symbiont (Brechenmacher et al. 2010; Libault et al. 2010a, c, d, e). These studies have led to the aforementioned essential gene *GmFWL1* (Libault et al. 2010c), as well as the identification of a soybean protein that has dual-localization in nodules (Libault et al. 2011). Transcript profiling with legume mutants and with plants on which nodule-like structures have been elicited have shown that there are discrete stages in the development of the symbioses as well as identified mutants with developmental defects in specific stages (Hogslund et al. 2009; Starker et al. 2006; Mitra and long 2004; Rightmyer and long 2011; Moreau et al. 2011). Laser microdissection of symbiotic tissues has revealed cell-type specific transcriptional responses (Guether et al. 2009; Hogekamp et al. 2011; Limpens et al. 2013).

Perspectives

The rapid development of high-throughput technologies allowing the characterization of genomic and transcriptomic, both protein coding and non-coding RNA sequences as well as the establishment of proteome and metabolome profiles is revolutionizing

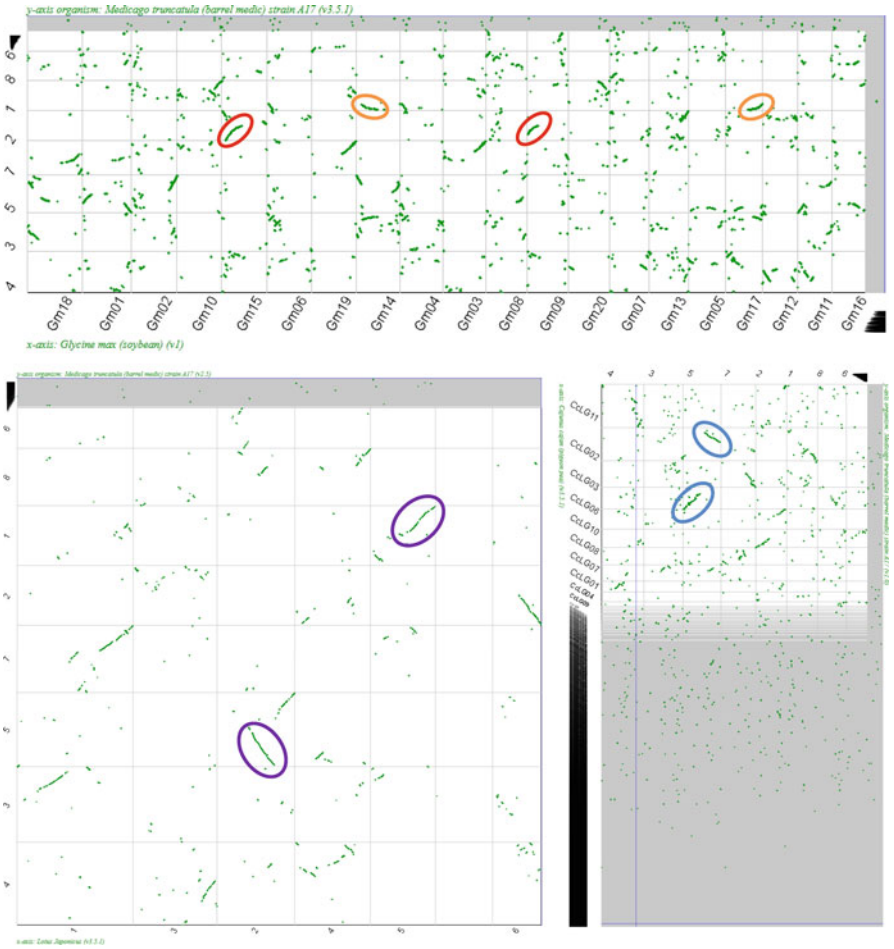


Fig. 2.2 Whole-genome dot-plot of between *M. truncatula* (Jemalong-A17) and various legume species: *G. max* (William82); *L. japonicus* (var. japonicus) and *Cajanus cajan* (pigeon pea). The x and y-axis *numbers* represent the chromosomes from each species. Each *dot* highlights a macro-synteny relationship between blocks of chromosome of the two species in comparision. Based on their recent evolutionary divergence, *M. truncatula* and *L. japonicus* chromosomes share strong syntenic relationships (*purple circles*) opposite to *M. truncatula* and *C. cajan* (*blue circles*). The recent duplication of the soybean genome is leading to a more complex interpretation of the synteny existing between *G. max* and *M. truncatula* (*orange* and *red circles*). Unanchored sequences are indicated in the *grey* section of each figure. This figure was generated using the SynMap tool available from the genome comparative platform CoGe (http://genomevolution.org/CoGe/)

legume functional genomics. Comparative genomic tools clearly highlight macrosyntenic relationships between non-model and model plants (Fig. 2.2). The synteny relationships existing between their genomes are directly correlated by the fact that the entire pool of legumes selected for genome sequencing belongs

to the Papilionoideae sub-family. Therefore, to enlarge our understanding of legume evolution, the generation of draft genomes from legumes selected from the other Mimosoideae and the Caesalpinoideae clades will be highly desirable. Among them, *Chamaecrista fasciculata*, a mimosoid, represents an attractive model to better understand legume evolution (Doyle 2011). First, in contrast to the four model legumes members which belong to the papilionoid sub-family, *C. fasciculata* genome did not undergo duplications events; thus *C. fasciculata* is expected to have a simpler genome structure. Second, in contrast to other mimosoids species sharing its clade, *C. fasciculata* is one of the nodulating mimosoids (Doyle 2011). Hence, it is possible that nodulation originated in *C. fasciculata* independently from model papilionoid legumes. Concomitant with this, translational genomics between legume genomes and the development of molecular tools allowing the deep investigation of gene function make legume functional genomics a very valuable strategy for improving legume biology. However, epigenetic regulation of gene expression in legumes, remain largely unexplored and will require immediate attention. A better understanding of legume evolution through the sequencing of non-model legumes including the genomes of related species can be expected to yield novel insights into legume biology that will be able to be translated to crop legume improvement.

References

Afzal AJ et al (2009) The nematode resistance allele at the rhg1 locus alters the proteome and primary metabolism of soybean roots. Plant Physiol 151(3):1264–1280

Agrawal GK et al (2008) In-depth investigation of the soybean seed-filling proteome and comparison with a parallel study of rapeseed. Plant Physiol 148(1):504–518

Alkharouf NW, Klink VP, Matthews BF (2007) Identification of *Heterodera glycines* (soybean cyst nematode [SCN]) cDNA sequences with high identity to those of *Caenorhabditis elegans* having lethal mutant or RNAi phenotypes. Exp Parasitol 115:247–258

Anderson JP et al (2010) The B-3 ethylene response factor MtERF1-1 mediates resistance to a subset of root pathogens in *Medicago truncatula* without adversely affecting symbiosis with rhizobia. Plant Physiol 154(2):861–873

Andriankaja A et al (2007) AP2-ERF transcription factors mediate Nod factor dependent Mt ENOD11 activation in root hairs via a novel cis-regulatory motif. Plant Cell 19(9):2866–2885

Aranjuelo I et al (2011) Plant physiology and proteomics reveals the leaf response to drought in alfalfa (*Medicago sativa* L.). J Exp Bot 62(1):111–123

Arenas-Huertero C et al (2009) Conserved and novel miRNAs in the legume *Phaseolus vulgaris* in response to stress. Plant Mol Biol 70(4):385–401

Ariel F et al (2010) Environmental regulation of lateral root emergence in *Medicago truncatula* requires the HD-Zip I transcription factor HB1. Plant Cell 22(7):2171–2183

Asamizu E et al (2000) Generation of 7137 non-redundant expressed sequence tags from a legume, *Lotus japonicus*. DNA Res 7:127–130

Asamizu E et al (2005) Comparison of the transcript profiles from the root and the nodulating root of the model legume *Lotus japonicus* by serial analysis of gene expression. Mol Plant Microbe Interact 18(5):487–498

Benedito VA et al (2008) A gene expression atlas of the model legume *Medicago truncatula*. Plant J 55(3):504–513

Benedito VA et al (2010) Genomic inventory and transcriptional analysis of *Medicago truncatula* transporters. Plant Physiol 152(3):1716–1730

Benlloch R et al (2006) Isolation of mtpim proves Tnt1 a useful reverse genetics tool in *Medicago truncatula* and uncovers new aspects of AP1-like functions in legumes. Plant Physiol 142: 972–983

Bolon YT et al (2011) Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. Plant Physiol 156(1):240–253

Boualem A et al (2008) MicroRNA166 controls root and nodule development in *Medicago truncatula*. Plant J 54(5):876–887

Bourgeois M et al (2009) Dissecting the proteome of pea mature seeds reveals the phenotypic plasticity of seed protein composition. Proteomics 9(2):254–271

Branscheid A et al (2010) Expression pattern suggests a role of MiR399 in the regulation of the cellular response to local Pi increase during arbuscular mycorrhizal symbiosis. Mol Plant Microbe Interact 23(7):915–926

Brechenmacher L et al (2008) Transcription profiling of soybean nodulation by *Bradyrhizobium japonicum*. Mol Plant Microbe Interact 21(5):631–645

Brechenmacher L et al (2009) Establishment of a protein reference map for soybean root hair cells. Plant Physiol 149(2):670–682

Brechenmacher L et al (2010) Soybean metabolites regulated in root hairs in response to the symbiotic bacterium *Bradyrhizobium japonicum*. Plant Physiol 153(4):1808–1822

Broughton W et al (2003) Beans (Phaseolus spp.)—model food legumes. Plant Soil 252:55–128

Buitink J et al (2006) Transcriptome profiling uncovers metabolic and regulatory processes occurring during the transition from desiccation-sensitive to desiccation-tolerant stages in *Medicago truncatula* seeds. Plant J 47(5):735–750

Caillaud MC et al (2008) Root-knot nematodes manipulate plant cell functions during a compatible interaction. J Plant Physiol 165(1):104–113

Caspi R et al (2010) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res 38(Database issue):D473–D479

Catoira R et al (2000) Four genes of *Medicago truncatula* controlling components of a Nod factor transduction pathway. Plant Cell 12(9):1647–1666

Chatelain E et al (2012) Temporal profiling of the heat-stable proteome during late maturation of *Medicago truncatula* seeds identifies a restricted subset of late embryogenesis abundant proteins associated with longevity. Plant Cell Environ 35(8):1440–1455

Chen R, Hu Z, Zhang H (2009) Identification of microRNAs in wild soybean (*Glycine soja*). J Integr Plant Biol 51(12):1071–1079

Chen JH et al (2010) Control of dissected leaf morphology by a Cys(2)His(2) zinc finger transcription factor in the model legume *Medicago truncatula*. Proc Natl Acad Sci U S A 107(23): 10754–10759

Chen ZJ et al (2011) Identification of differentially expressed proteins in soybean nodules under phosphorus deficiency through proteomic analysis. Proteomics 11(24):4648–4659

Chen L et al (2012a) Ethylene-responsive miRNAs in roots of *Medicago truncatula* identified by high-throughput sequencing at whole genome level. Plant Sci 184:14–19

Chen L et al (2012b) Identification of aluminum-responsive microRNAs in *Medicago truncatula* by genome-wide high-throughput sequencing. Planta 235(2):375–386

Cheng X et al (2011) Reverse genetics in *Medicago truncatula* using Tnt1 insertion mutants. Methods Mol Biol 678:179–190

Choi JJ et al (2008) Expression patterns in soybean resistant to *Phakopsora pachyrhizi* reveal the importance of peroxidases and lipoxygenases. Funct Integr Genomics 8(4):341–359

Colbert T et al (2001) High-throughput screening for induced point mutations. Plant Physiol 126(2):480–484

Colebatch G et al (2002) Novel aspects of symbiotic nitrogen fixation uncovered by transcript profiling with cDNA arrays. Mol Plant Microbe Interact 15(5):411–420

Colebatch G et al (2004) Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus japonicus*. Plant J 39(4):487–512

Combier JP et al (2006) MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. Genes Dev 20(22):3084–3088

Cook DR (1999) *Medicago truncatula*—a model in the making! Curr Opin Plant Biol 2:301–304
Cooper JL et al (2008) TILLING to detect induced mutations in soybean. BMC Plant Biol 8:9
Cortes AJ, Chavarro MC, Blair MW (2011) SNP marker diversity in common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 123(5):827–845
Covitz PA, Smith LS, Long SR (1998) Expressed sequence tags from a root-hair-enriched *Medicago truncatula* cDNA library. Plant Physiol 117:1325–1332
Dam S et al (2009) The proteome of seed development in the model legume *Lotus japonicus*. Plant Physiol 149(3):1325–1340
Das S et al (2010) Transcriptional profiling of root-knot nematode induced feeding sites in cowpea (*Vigna unguiculata* L. Walp.) using a soybean genome array. BMC Genomics 11:16
Davis EL, Mitchum MG (2005) Nematodes. Sophisticated parasites of legumes. Plant Physiol 137(4):1182–1188
Deguchi Y et al (2007) Transcriptome profiling of *Lotus japonicus* roots during arbuscular mycorrhiza development and comparison with that of nodulation. DNA Res 14:117–133
Devers EA et al (2011) Stars and symbiosis: microRNA- and microRNA*-mediated transcript cleavage involved in arbuscular mycorrhizal symbiosis. Plant Physiol 156(4):1990–2010
Dierking EC, Bilyeu KD (2009) New sources of soybean seed meal and oil composition traits identified through TILLING. BMC Plant Biol 9:89
Doyle JJ (2011) Phylogenetic perspectives on the origins of nodulation. Mol Plant Microbe Interact 24(11):1289–1295
El Yahyaoui F et al (2004) Expression profiling in *Medicago truncatula* identifies more than 750 genes differentially expressed during nodulation, including many potential regulators of the symbiotic program. Plant Physiol 136(2):3159–3176
Endo M et al (2000) Analysis of expressed sequence tags of flower buds in *Lotus japonicus*. DNA Res 7:213–216
Farag MA et al (2008) Metabolomics reveals novel pathways and differential mechanistic and elicitor-specific responses in phenylpropanoid and isoflavonoid biosynthesis in *Medicago truncatula* cell cultures. Plant Physiol 146(2):387–402
Fedorova M et al (2002) Genome-wide identification of nodule-specific transcripts in the model legume *Medicago truncatula*. Plant Physiol 130(2):519–537
Frenzel A et al (2005) Combined transcriptome profiling reveals a novel family of arbuscular mycorrhizal-specific *Medicago truncatula* lectin genes. Mol Plant Microbe Interact 18(8): 771–782
Gaige AR, Ayella A, Shuai B (2011) Methyl jasmonate and ethylene induce partial resistance in *Medicago truncatula* against the charcoal rot pathogen Macrophomina phaseolina. Physiol Mol Plant Pathol 74:412–418
Gallardo K et al (2003) Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. Plant Physiol 133(2): 664–682
Gallardo K et al (2007) A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds: evidence for metabolic specialization of maternal and filial tissues. Mol Cell Proteomics 6(12):2165–2179
Gao LL et al (2010) Identification of potential early regulators of aphid resistance in *Medicago truncatula* via transcription factor expression profiling. New Phytol 186(4):980–994
Garcia A et al (2008) Molecular mapping of soybean rust (*Phakopsora pachyrhizi*) resistance genes: discovery of a novel locus and alleles. Theor Appl Genet 117(4):545–553
Garg R et al (2011) Gene discovery and tissue-specific transcriptome analysis in chickpea with massively parallel pyrosequencing and web resource development. Plant Physiol 156(4):1661–1678
Gomez SK et al (2009) *Medicago truncatula* and *Glomus intraradices* gene expression in cortical cells harboring arbuscules in the arbuscular mycorrhizal symbiosis. BMC Plant Biol 9:10
Graham M et al (2006) Identification of candidate phosphorus stress induced genes in *Phaseolus vulgaris* through clustering analysis across several plant species. Funct Plant Biol 33:789–797
Gruber V et al (2009) Identification of transcription factors involved in root apex responses to salt stress in *Medicago truncatula*. Mol Genet Genomics 281(1):55–66

Guether M et al (2009) Genome-wide reprogramming of regulatory networks, transport, cell wall and membrane biogenesis during arbuscular mycorrhizal symbiosis in *Lotus japonicus*. New Phytol 182:200–212

Gyorgyey J et al (2000) Analysis of *Medicago truncatula* nodule expressed sequence tags. Mol Plant Microbe Interact 13:62–71

Hajduch M et al (2006) Proteomic analysis of seed filling in *Brassica napus.* Developmental characterization of metabolic isozymes using high-resolution two-dimensional gel electrophoresis. Plant Physiol 141(1):32–46

Han Y et al (2011) Genome-wide SNP discovery in tetraploid alfalfa using 454 sequencing and high resolution melting analysis. BMC Genomics 12:1–11

Hancock CN et al (2011) The rice miniature inverted repeat transposable element mPing is an effective insertional mutagen in soybean. Plant Physiol 157(2):552–562

He J et al (2009) The *Medicago truncatula* gene expression atlas web server. BMC Bioinformatics 10:441

Hernandez G et al (2007) Phosphorus stress in common bean: root transcript and metabolic responses. Plant Physiol 144(2):752–767

Hernandez G et al (2009) Global changes in the transcript and metabolic profiles during symbiotic nitrogen fixation in phosphorus-stressed common bean plants. Plant Physiol 151(3):1221–1238

Hiremath PJ et al (2011) Large-scale transcriptome analysis in chickpea (*Cicer arietinum* L.), an orphan legume crop of the semi-arid tropics of Asia and Africa. Plant Biotechnol J 9(8): 922–931

Hirsch S et al (2009) GRAS proteins form a DNA binding complex to induce gene expression during nodulation signaling in *Medicago truncatula*. Plant Cell 21(2):545–557

Hoffmann D et al (2007) Nodulation deficiency caused by fast neutron mutagenesis of the model legume *Lotus japonicus*. J Plant Physiol 164(4):460–469

Hogekamp C et al (2011) Laser microdissection unravels cell-type-specific transcription in arbuscular mycorrhizal roots, including CAAT-box transcription factor gene expression correlating with fungal contact and spread. Plant Physiol 157(4):2023–2043

Hogslund N et al (2009) Dissection of symbiosis and organ development by integrated transcriptome analysis of *Lotus japonicus* mutant and wild-type plants. PLoS One 4(8):e6556

Hohnjec N et al (2005) Overlaps in the transcriptional profiles of *Medicago truncatula* roots inoculated with two different Glomus fungi provide insights into the genetic program activated during arbuscular mycorrhiza. Plant Physiol 137:1283–1301

Holmes P et al (2008) Transcriptional profiling of *Medicago truncatula* meristematic root cells. BMC Plant Biol 8:21

Hruz T et al (2008) Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. Adv Bioinformatics 2008:420747

Hyten DL et al (2010) High-throughput SNP discovery and assay development in common bean. BMC Genomics 11:475

Ibrahim HMM et al (2011) Analysis of gene expression in soybean (*Glycine max*) roots in response to the root knot nematode Meloidogyne incognita using microarrays and KEGG pathways. BMC Genomics 12:16

Imin N et al (2008) Genome-wide transcriptional analysis of super-embryogenic *Medicago truncatula* explant cultures. BMC Plant Biol 8:110

Iqbal MJ et al (2005) Root response to *Fusarium solani* f. sp. *glycines*: temporal accumulation of transcripts in partially resistant and susceptible soybean. Theor Appl Genet 110:1429–1438

Ithal N et al (2007a) Developmental transcript profiling of cyst nematode feeding cells in soybean roots. Mol Plant Microbe Interact 20(5):510–525

Ithal N et al (2007b) Parallel genome-wide expression profiling of host and pathogen during soybean cyst nematode infection of soybean. Mol Plant Microbe Interact 20(3):293–305

Jagadeeswaran G et al (2009) Cloning and characterization of small RNAs from *Medicago truncatula* reveals four novel legume-specific microRNA families. New Phytol 184(1):85–98

Jasinski M et al (2009) Changes in the profile of flavonoid accumulation in *Medicago truncatula* leaves during infection with fungal pathogen *Phoma medicaginis*. Plant Physiol Biochem 47(9):847–853

Jones KM et al (2008) Differential response of the plant *Medicago truncatula* to its symbiont *Sinorhizobium meliloti* or an exopolysaccharide-deficient mutant. Proc Natl Acad Sci U S A 105(2):704–709

Joshi T et al (2010) Prediction of novel miRNAs and associated target genes in *Glycine max*. BMC Bioinformatics 11(Suppl 1):S14

Joshi T et al (2012) Soybean Knowledge Base (SoyKB): a web resource for soybean translational genomics. BMC Genomics 13(Suppl 1):S15

Journet EP et al (2002) Exploring root symbiotic programs in the model legume *Medicago truncatula* using EST analysis. Nucleic Acids Res 30(24):5579–5592

Kakar K et al (2008) A community resource for high-throughput quantitative RT-PCR analysis of transcription factor gene expression in *Medicago truncatula*. Plant Methods 4:18

Kalavacharla V et al (2011) Identification and analysis of common bean (*Phaseolus vulgaris* L.) transcriptomes by massively parallel pyrosequencing. BMC Plant Biol 11:135

Kalo P et al (2005) Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. Science 308(5729):1786–1789

Kamphuis LG et al (2010) Phoma medicaginis stimulates the induction of the octadecanoid and phenylpropanoid pathways in *Medicago truncatula*. Mol Plant Pathol. doi:10.1111/j.1364-3703.2011.00767.x

Kandoth PK et al (2011) The soybean Rhg1 locus for resistance to the soybean cyst nematode *Heterodera glycines* regulates the expression of a large number of stress- and defense-related genes in degenerating feeding cells. Plant Physiol 155(4):1960–1975

Kang Y et al (2011) System responses to long-term drought and re-watering of two contrasting alfalfa varieties. Plant J 68(5):871–889

Klink VP, Matthews BF (2009) Emerging approaches to broaden resistance of soybean to soybean cyst nematode as supported by gene expression studies. Plant Physiol 151(3):1017–1022

Klink VP et al (2007a) A time-course comparative microarray analysis of an incompatible and compatible response by *Glycine max* (soybean) to *Heterodera glycines* (soybean cyst nematode) infection. Planta 226(6):1423–1447

Klink VP et al (2007b) Laser capture microdissection (LCM) and comparative microarray expression analysis of syncytial cells isolated from incompatible and compatible soybean (*Glycine max*) roots infected by the soybean cyst nematode (*Heterodera glycines*). Planta 226(6): 1389–1409

Klink VP et al (2009a) A correlation between host-mediated expression of parasite genes as tandem inverted repeats and abrogation of development of female *Heterodera glycines* cyst formation during infection of *Glycine max*. Planta 230(1):53–71

Klink VP et al (2009b) A gene expression analysis of syncytia laser microdissected from the roots of the *Glycine max* (soybean) genotype PI 548402 (Peking) undergoing a resistant reaction after infection by *Heterodera glycines* (soybean cyst nematode). Plant Mol Biol 71(6): 525–567

Klink VP et al (2010a) Syncytium gene expression in *Glycine max*([PI 88788]) roots undergoing a resistant reaction to the parasitic nematode Heterodera glycines. Plant Physiol Biochem 48(2–3):176–193

Klink VP et al (2010b) The application of a developmental genomics approach to study the resistant reaction of soybean to the soybean cyst nematode. Nematropica 40(1):1–11

Klink VP et al (2011) Differences in gene expression amplitude overlie a conserved transcriptomic program occurring between the rapid and potent localized resistant reaction at the syncytium of the *Glycine max* genotype Peking (PI 548402) as compared to the prolonged and potent resistant reaction of PI 88788. Plant Mol Biol 75(1–2):141–165

Kouchi H et al (2004) Large-scale analysis of gene expression profiles during early stages of root nodule formation in a model legume, *Lotus japonicus*. DNA Res 11(4):263–274

Krishnan HB, Oehrle NW, Natarajan SS (2009) A rapid and simple procedure for the depletion of abundant storage proteins from legume seeds to advance proteome analysis: a case study using *Glycine max*. Proteomics 9(11):3174–3188

Kulcheski FR et al (2011) Identification of novel soybean microRNAs involved in abiotic and biotic stresses. BMC Genomics 12:307
Kulinski J et al (2000) CEL I enzymatic mutation detection assay. Biotechniques 29(1):44–46, 48
Kuster H et al (2004) Construction and validation of cDNA-based Mt6k-RIT macro- and microarrays to explore root endosymbioses in the model legume *Medicago truncatula*. J Biotechnol 108(2):95–113
Le BH et al (2007) Using genomics to study legume seed development. Plant Physiol Biochem 144:562–574
Le Roux MR et al (2006) Routes of pyruvate synthesis in phosphorus-deficient lupin roots and nodules. New Phytol 169(2):399–408
Le Signor C et al (2009) Optimizing TILLING populations for reverse genetics in *Medicago truncatula*. Plant Biotechnol J 7(5):430–441
Lee Y et al (2005) The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. Nucleic Acids Res 33:D71–D74
Lelandais-Brière C et al (2009) Genome-wide Medicago truncatula small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. Plant Cell 21:2780–2796
Li H et al (2010) Misexpression of miR482, miR1512, and miR1515 increases soybean nodulation. Plant Physiol 153(4):1759–1770
Libault M et al (2009) Large-scale analysis of putative soybean regulatory gene expression identifies a Myb gene involved in soybean nodule development. Plant Physiol 151(3):1207–1220
Libault M et al (2010a) A member of the highly conserved FWL (tomato FW2.2-like) gene family is essential for soybean nodule organogenesis. Plant J 62(5):852–864
Libault M et al (2010b) Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. Plant Physiol 152(2): 541–552
Libault M et al (2010c) Root hair systems biology. Trends Plant Sci 15(11):641–650
Libault M et al (2010d) An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. Plant J 63(1):86–99
Libault M et al (2010e) Soybean root hairs: a valuable system to investigate plant biology at the cellular level. Plant Signal Behav 5(4):419–421
Libault M et al (2011) A dual-targeted soybean protein is involved in *Bradyrhizobium japonicum* infection of soybean root hair and cortical cells. Mol Plant Microbe Interact 24:1051–1060
Limpens E et al (2013) Cell- and tissue-specific transcriptome analyses of *Medicago truncatula* root nodules. PLoS One 8(5):e64377. doi:10.1371/journal.pone.0064377
Liu J et al (2003) Transcript profiling coupled with spatial expression analyses reveals genes involved in distinct developmental stages of an arbuscular mycorrhizal symbiosis. Plant Cell 15:2106–2123
Liu J et al (2007) Arbuscular mycorrhizal symbiosis is accompanied by local and systemic alterations in gene expression and an increase in disease resistance in the shoots. Plant J 50: 529–544
Lohar DP, Bird DM (2003) *Lotus japonicus*: a new model to study root-parasitic nematodes. Plant Cell Physiol 44(11):1176–1184
Lohar DP et al (2006) Transcript analysis of early nodulation events in *Medicago truncatula*. Plant Physiol 140(1):221–234
Madrid E et al (2010) Transcription factor profiling leading to the identification of putative transcription factors involved in the *Medicago truncatula*-Uromyces striatus interaction. Theor Appl Genet 121(7):1311–1321
Manthey K et al (2004) Transcriptome profiling in root nodules and arbuscular mycorrhiza identifies a collection of novel genes induced during *Medicago truncatula* root endosymbioses. Mol Plant Microbe Interact 17(10):1063–1077
Margam VM et al (2011) Transcriptome sequencing, and rapid development and application of SNP markers for the legume pod borer *Maruca vitrata* (Lepidoptera: Crambidae). PLoS One 6(7):e21388

Marsolais F et al (2010) Proteomic analysis of common bean seed with storage protein deficiency reveals up-regulation of sulfur-rich proteins and starch and raffinose metabolic enzymes, and down-regulation of the secretory pathway. J Proteomics 73(8):1587–1600
Matsye PD et al (2011) Mapping cell fate decisions that occur during soybean defense responses. Plant Mol Biol 77(4–5):513–528
Mazarei M et al (2011) Gene expression profiling of resistant and susceptible soybean lines infected with soybean cyst nematode. Theor Appl Genet 123(7):1193–1206
Middleton PH et al (2007) An ERF transcription factor in *Medicago truncatula* that is essential for Nod factor signal transduction. Plant Cell 19:1221–1234
Mitra RM, Long SR (2004) Plant and bacterial symbiotic mutants define three transcriptionally distinct stages in the development of the *Medicago truncatula*/Sinorhizobium meliloti symbiosis. Plant Physiol 134(2):595–604
Mitra RM et al (2004a) A Ca2+/calmodulin-dependent protein kinase required for symbiotic nodule development: gene identification by transcript-based cloning. Proc Natl Acad Sci U S A 101(13):4701–4705
Mitra RM, Shaw SL, Long SR (2004b) Six nonnodulating plant mutants defective for Nod factor-induced transcriptional changes associated with the legume-rhizobia symbiosis. Proc Natl Acad Sci U S A 101(27):10217–10222
Moreau S et al (2011) Transcription reprogramming during root nodule development in *Medicago truncatula*. PLoS One 6(1):e16463
Muchero W et al (2011) Genic SNP markers and legume synteny reveal candidate genes underlying QTL for Macrophomina phaseolina resistance and maturity in cowpea [*Vigna unguiculata* (L) Walp.]. BMC Genomics 12:8
Mueller LA, Zhang P, Rhee SY (2003) AraCyc: a biochemical pathway database for Arabidopsis. Plant Physiol 132(2):453–460
Muino JM, Angenent GC, Kaufmann K (2011) Visualizing and characterizing in vivo DNA-binding events and direct target genes of plant transcription factors. Methods Mol Biol 754:293–305
Murray JD et al (2011) Vapyrin, a gene essential for intracellular progression of arbuscular mycorrhizal symbiosis, is also essential for infection by rhizobia in the nodule symbiosis of *Medicago truncatula*. Plant J 65(2):244–252
Naoumkina M et al (2007) Different mechanisms for phytoalexin induction by pathogen and wound signals in *Medicago truncatula*. Proc Natl Acad Sci U S A 104(46):17909–17915
Naoumkina M et al (2008) Metabolic and genetic perturbations accompany the modification of galactomannan in seeds of *Medicago truncatula* expressing mannan synthase from guar (*Cyamopsis tetragonoloba* L.). Plant Biotechnol J 6(6):619–631
Nautrup-Pedersen G et al (2010) Proteome analysis of pod and seed development in the model legume *Lotus japonicus*. J Proteome Res 9(11):5715–5726
Neelakandan AK et al (2012) Metabolic engineering of soybean affords improved phytosterol seed traits. Plant Biotechnol J 10:12–19
Olivera M et al (2004) Growth, nitrogen fixation and ammonium assimilation in common bean (*Phaseolus vulgaris*): effect of phosphorus. Physiol Plant 121(3):498–505
Omrane S et al (2009) Symbiotic competence in *Lotus japonicus* is affected by plant nitrogen status: transcriptomic identification of genes affected by a new signalling pathway. New Phytol 183:380–394
Pang Y et al (2008) A transcript profiling approach reveals an epicatechin-specific glucosyltransferase expressed in the seed coat of *Medicago truncatula*. Proc Natl Acad Sci U S A 105(37):14210–14215
Panthee DR et al (2009) Differential expression of genes in soybean in response to the causal agent of Asian soybean rust (*Phakopsora pachyrhizi* Sydow) is soybean growth stage-specific. Theor Appl Genet 118(2):359–370
Payton P et al (2009) Gene expression profiling in peanut using high density oligonucleotide microarrays. BMC Genomics 10:265

Penmetsa RV, Cook DR (2000) Production and characterization of diverse developmental mutants of *Medicago truncatula*. Plant Physiol 123(4):1387–1398
Perry JA et al (2003) A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. Plant Physiol 131(3):866–871
Perry J et al (2009) TILLING in *Lotus japonicus* identified large allelic series for symbiosis genes and revealed a bias in functionally defective ethyl methanesulfonate alleles toward glycine replacements. Plant Physiol 151(3):1281–1291
Poch HLC, Lopez RHM, Clark SJ (2007) Ecotypes of the model legume *Lotus japonicus* vary in their interaction phenotypes with the root-knot nematode Meloidogyne incognita. Ann Bot 99(6):1223–1229
Porceddu A et al (2008) An Italian functional genomic resource for *Medicago truncatula*. BMC Res Notes 1:129
Puthoff DP et al (2007) GeneChip profiling of transcriptional responses to soybean cyst nematode, Heterodera glycines, colonization of soybean roots. J Exp Bot 58(12):3407–3418
Quesenberry KH et al (2008) Response of native southeastern US legumes to root-knot nematodes. Crop Sci 48(6):2274–2278
Radwan O, Liu Y, Clough SJ (2011) Transcriptional analysis of soybean root response to Fusarium virguliforme, the causal agent of sudden death syndrome. Mol Plant Microbe Interact 24(8):958–972
Rakocevic A et al (2009) MERE1, a low-copy-number copia-type retroelement in *Medicago truncatula* active during tissue culture. Plant Physiol 151:1250–1263
Ramírez-Suero M et al (2010) A study on the susceptibility of the model legume plant *Medicago truncatula* to the soil-borne pathogen *Fusarium oxysporum*. Eur J Plant Pathol 126:517–530
Repetto O et al (2008) Exploring the nuclear proteome of *Medicago truncatula* at the switch towards seed filling. Plant J 56(3):398–410
Ribet J, Drevon JJ (1995) Phosphorus deficiency increases the acetylene-induced decline in nitrogenase activity in soybean (*Glycine-Max* (L) Merr). J Exp Bot 46(291):1479–1486
Rightmyer AP, Long SR (2011) Pseudonodule formation by wild-type and symbiotic mutant *Medicago truncatula* in response to auxin transport inhibitors. Mol Plant Microbe Interact 24(11):1372–1384
Rogers C et al (2009) Deletion-based reverse genetics in *Medicago truncatula*. Plant Physiol 151:1077–1086
Ruffel S et al (2008) Systemic signaling of the plant nitrogen status triggers specific transcriptome responses depending on the nitrogen source in *Medicago truncatula*. Plant Physiol 146(4):2020–2035
Sa TM, Israel DW (1991) Energy status and functioning of phosphorus-deficient soybean nodules. Plant Physiol 97:928–935
Sagan M et al (1995) Selection of nodulation and mycorrhizal mutants in the model plant *Medicago truncatula* (Gaertn.) after g-mutagenesis. Plant Sci 111:63–71
Sakata K et al (2009) Soybean proteome database: a data resource for plant differential omics. J Proteome Res 8(7):3539–3548
Samac DA, Graham MA (2007) Recent advances in legume-microbe interactions: recognition, defense response, and symbiosis from a genomic perspective. Plant Physiol 144(2):582–587
Sanchez DH et al (2008a) Integrative functional genomics of salt acclimatization in the model legume *Lotus japonicus*. Plant J 53(6):973–987
Sanchez DH et al (2008b) Plant metabolomics reveals conserved and divergent metabolic responses to salinity. Physiol Plant 132(2):209–219
Sanchez DH et al (2010) Mining for robust transcriptional and metabolic responses to long-term salt stress: a case study on the model legume *Lotus japonicus*. Plant Cell Environ 33(4):468–480
Sanchez DH et al (2011) Comparative functional genomics of salt stress in related model and cultivated plants identifies and overcomes limitations to translational genomics. PLoS One 6(2):11
Sanchez DH et al (2012) Comparative metabolomics of drought acclimation in model and forage legumes. Plant Cell Environ 35(1):136–149

Sawbridge T et al (2003) Generation and analysis of expressed sequence tags in white clover (*Trifolium repens* L.). Plant Sci 165:1077–1087
Schauser L et al (1999) A plant regulator controlling development of symbiotic root nodules. Nature 402(6758):191–195
Schmitz RJ, Zhang X (2011) High-throughput approaches for plant epigenomic studies. Curr Opin Plant Biol 14(2):130–136
Schmutz J et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463(7278): 178–183
Schneider KT et al (2011) Biphasic gene expression changes elicited by *Phakopsora pachyrhizi* in soybean correlate with fungal penetration and haustoria formation. Plant Physiol 157(1): 355–371
Schroeyers K et al (2004) Nodulation-enhanced sequences from the water stress-tolerant tropical legume Sesbania rostrata. Plant Sci 167:207–216
Severin AJ et al (2010) RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. BMC Plant Biol 10:160
Shoemaker R et al (2002) A compilation of soybean ESTs: generation and analysis. Genome 45:329–338
Shu Y et al (2011) SNPs discovery and CAPS marker conversion in soybean. Mol Biol Rep 38(3):1841–1846
Simon SA, Meyers BC, Sherrier DJ (2009) MicroRNAs in the rhizobia legume symbiosis. Plant Physiol 151:1002–1008
Song QX et al (2011) Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. BMC Plant Biol 11:5
Soria-Guerra RE et al (2010a) Identifying differentially expressed genes in leaves of Glycine tomentella in the presence of the fungal pathogen *Phakopsora pachyrhizi*. Planta 232(5):1181–1189
Soria-Guerra RE et al (2010b) Transcriptome analysis of resistant and susceptible genotypes of Glycine tomentella during *Phakopsora pachyrhizi* infection reveals novel rust resistance genes. Theor Appl Genet 120(7):1315–1333
Starker CG et al (2006) Nitrogen fixation mutants of *Medicago truncatula* fail to support plant and bacterial symbiotic gene expression. Plant Physiol 140(2):671–680
Stirling GR et al (2006) The reaction of soybean and other legume crops to root-knot nematodes (Meloidogyne spp.), and implications for growing these crops in rotation with sugarcane. Australas Plant Pathol 35(6):707–714
Subramanian S et al (2008) Novel and nodulation-regulated microRNAs in soybean roots. BMC Genomics 9:160
Suganuma N et al (2004) CDNA macroarray analysis of gene expression in ineffective nodules induced on the *Lotus japonicus* sen1 mutant. Mol Plant Microbe Interact 17(11):1223–1233
Sunkar R, Jagadeeswaran G (2008) In silico identification of conserved microRNAs in large number of diverse plant species. BMC Plant Biol 8:37
Szczyglowski K et al (1997) Construction of a *Lotus japonicus* late nodulin expressed tab library and identification of novel nodule-specific genes. Plant Physiol 114:1335–1346
Szittya G et al (2008) High-throughput sequencing of *Medicago truncatula* short RNAs identifies eight new miRNA families. BMC Genomics 9:593
Tadege M, Ratet P, Mysore KS (2005) Insertional mutagenesis: a Swiss Army knife for functional genomics of *Medicago truncatula*. Trends Plant Sci 10(5):229–235
Tadege M et al (2008) Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume *Medicago truncatula*. Plant J 54(2):335–347
Takanashi K et al (2012) Tissue-specific transcriptome analysis in nodules of *Lotus japonicus*. Plant Physiol 159(1):501–516
Tang C et al (2001) Phosphorus deficiency impairs early nodule functioning and enhances proton release in roots of *Medicago truncatula* L. Ann Bot 88(1):131–138
Tremblay A et al (2010) Transcriptome analysis of a compatible response by *Glycine max* to *Phakopsora pachyrhizi* infection. Plant Sci 179(3):183–193

Trindade I et al (2010) miR398 and miR408 are up-regulated in response to water deficit in *Medicago truncatula*. Planta 231(3):705–716
Udvardi MK et al (2007) Legume transcription factors: global regulators of plant development and response to the environment. Plant Physiol 144(2):538–549
Uppalapati SR et al (2009) Global gene expression profiling during *Medicago truncatula-Phymatotrichopsis omnivora* interaction reveals a role for jasmonic acid, ethylene, and the flavonoid pathway in disease development. Mol Plant Microbe Interact 22(1):7–17
Uppalapati SR et al (2010) Phymatotrichum (cotton) root rot caused by *Phymatotrichopsis omnivora*: retrospects and prospects. Mol Plant Pathol 11(3):325–334
Uppalapati SR et al (2012) Loss of abaxial leaf epicuticular wax in Medicago truncatula irg1/palm1 mutants results in reduced spore differentiation of anthracnose and nonhost rust pathogens. Plant Cell 24(1):353–370. doi:10.1105/tpc.111.093104. Epub 2012 Jan 31
Urbanczyk-Wochniak E, Sumner LW (2007) MedicCyc: a biochemical pathway database for *Medicago truncatula*. Bioinformatics 23(11):1418–1423
Valachovicova T et al (2004) Soy isoflavones suppress invasiveness of breast cancer cells by the inhibition of NF-kappaB/AP-1-dependent and -independent pathways. Int J Oncol 25(5): 1389–1395
Valdes-Lopez O et al (2008) Essential role of MYB transcription factor: PvPHR1 and microRNA: PvmiR399 in phosphorus-deficiency signalling in common bean roots. Plant Cell Environ 31(12):1834–1843
Valdes-Lopez O et al (2010) MicroRNA expression profile in common bean (*Phaseolus vulgaris*) under nutrient deficiency stresses and manganese toxicity. New Phytol 187(3):805–818
Valdes-Lopez O et al (2011) Identification of quantitative trait loci controlling gene expression during the innate immunity response of soybean. Plant Physiol 157(4):1975–1986
van de Mortel M et al (2007) Distinct biphasic mRNA changes in response to Asian soybean rust infection. Mol Plant Microbe Interact 20(8):887–899
Vance CP, Uhde-Stone C, Allan DL (2003) Phosphorus acquisition and use: critical adaptations by plants for securing a nonrenewable resource. New Phytol 157(3):423–447
Varala K et al (2011) Rapid genotyping of soybean cultivars using high throughput sequencing. PLoS One 6(9):e24811
Verdier J et al (2008) Gene expression profiling of M. truncatula transcription factors identifies putative regulators of grain legume seed filling. Plant Mol Biol 67(6):567–580
Verdier J et al (2012) MtPAR MYB transcription factor acts as an on switch for proanthocyanidin biosynthesis in *Medicago truncatula*. Proc Natl Acad Sci U S A 109:109
Vernie T et al (2008) EFD is an ERF transcription factor involved in the control of nodule number and differentiation in *Medicago truncatula*. Plant Cell 20(10):2696–2713
Vodkin LO et al (2004) Microarrays for global expression constructed with a low redundancy set of 27,500 sequenced cDNAs representing an array of developmental stages and physiological conditions of the soybean plant. BMC Genomics 5:73
Wang M et al (2010) Inferring large-scale gene regulatory networks using a low-order constraint-based algorithm. Mol Biosyst 6(6):988–998
Wang T et al (2011) Identification of drought-responsive microRNAs in *Medicago truncatula* by genome-wide high-throughput sequencing. BMC Genomics 12:367
Wang XD et al (2012) From embryo sac to oil and protein bodies: embryo development in the model legume *Medicago truncatula*. New Phytol 193:327–338
Wasson AP et al (2009) Differing requirements for flavonoids during the formation of lateral roots, nodules and root knot nematode galls in *Medicago truncatula*. New Phytol 183(1):167–179
Watson BS et al (2003) Mapping the proteome of barrel medic (*Medicago truncatula*). Plant Physiol 131(3):1104–1123
Wise RP et al (2007) Transcript profiling in host-pathogen interactions. Annu Rev Phytopathol 45:329–369
Wong MM, Cannon CH, Wickneswari R (2011) Identification of lignin genes and regulatory sequences involved in secondary cell wall formation in Acacia auriculiformis and Acacia mangium via de novo transcriptome sequencing. BMC Genomics 12:342

Wu W et al (2008) Comparative metabolic profiling reveals secondary metabolites correlated with soybean salt tolerance. J Agric Food Chem 56(23):11132–11138
Wu T et al (2009) Mass spectrometry analysis of the variants of histone H3 and H4 of soybean and their post-translational modifications. BMC Plant Biol 9:98
Xu P et al (2011) A SNP and SSR based genetic map of asparagus bean (*Vigna. unguiculata* ssp. sesquipedialis) and comparison with the broader species. PLoS One 6(1):e15952
Xuan P et al (2011) MaturePred: efficient identification of microRNAs within novel plant pre-miRNAs. PLoS One 6(11):e27422
Yang SS et al (2010) Transcript profiling of common bean (*Phaseolus vulgaris* L.) using the GeneChip Soybean Genome Array: optimizing analysis by masking biased probes. BMC Plant Biol 10:85
You J et al (2011) Transcriptomic responses to aluminum stress in soybean roots. Genome 54(11):923–933
Young ND et al (2011) The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature 480(7378):520–524
Zahaf O et al (2012) Comparative transcriptomic analysis of salt adaptation in roots of contrasting *Medicago truncatula* genotypes. Mol Plant 5(5):1068–1081
Zhai J et al (2011) MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. Genes Dev 25:2540–2553
Zhang K et al (2006) The *Medicago truncatula* small protein proteome and peptidome. J Proteome Res 5(12):3355–3367
Zhang B, Pan X, Stellwag EJ (2008) Identification of soybean microRNAs and their targets. Planta 229(1):161–182
Zhang Z et al (2010) PMRD: plant microRNA database. Nucleic Acids Res 38(Database issue):D806–D813
Zhao CZ et al (2010) Deep sequencing identifies novel and conserved microRNAs in peanuts (*Arachis hypogaea* L.). BMC Plant Biol 10:3
Zhe Y et al (2013) miR172 regulates soybean nodulation. Mol Plant Microbe Interact. PMID: 23980625

Chapter 3
Advances in Soybean Genomics

Milind B. Ratnaparkhe, S.V. Ramesh, Giriraj Kumawat, S.M. Husain, and Sanjay Gupta

Abstract Soybean is an agronomically important crop that is endowed with rich seed protein and oil. It enriches the soil by fixing nitrogen through symbiosis with bacteria. In addition to human consumption, soybean is a major protein source in animal feeds and is also becoming a major crop for biodiesel production. A major landmark in soybean genomics research was its draft genome sequence assembly (cultivar Williams 82) following whole-genome shot gun (WGS) approach. It revealed 950 Mb (megabases) of assembled and anchored sequence as against the predicted 1,115 Mb genome consequently representing 85 % of the whole genome. Development of comprehensive physical map employing chiefly Bacterial artificial chromosomes (BAC) and Binary large-insert BAC clones (BIBAC) have assisted in the whole genome sequencing venture and in targeted genetic marker development, accelerating positional cloning approaches along with the generation of rapid and robust EST maps. Comprehensive Expressed Sequence Tags (ESTs) repository and genome sequence of the crop have helped in sound integration of physical map with the genetic map. In order to perform genetic and genomic analysis various molecular markers like RFLP, RAPD, AFLP, SSR, SNP etc. have been employed on RIL or F2 populations. In addition the genome is typified with single nucleotide polymorphisms (SNPs) and its utilization in molecular breeding applications like QTL mapping, positional cloning and association mapping studies is gaining impetus. QTLs associated with foremost traits of agronomic interests including QTLs for Aphid resistance, Soybean Cyst Nematode (SCN) resistance among others have been identified and validated. Further molecular marker assisted QTL introgression and gene

M.B. Ratnaparkhe, Ph.D. • S.V. Ramesh, Ph.D. • G. Kumawat, Ph.D.
S.M. Husain, Ph.D. • S. Gupta, Ph.D. (✉)
Crop Improvement Section, Directorate of Soybean Research,
Khandwa Road, Indore, MP, India
e-mail: sanitaishu@gmail.com

S. Gupta et al. (eds.), *Legumes in the Omic Era*, DOI 10.1007/978-1-4614-8370-0_3,

pyramiding for traits like enhanced seed protein concentration and Soybean Mosaic Virus (SMV) resistance, insect resistance etc. have been accomplished. Legume comparative genomics using orthologous genomic regions have addressed queries relating to Nucleotide binding-Leucine rich repeat (NB-LRRs) genes, polyploidy, and genome evolution. In the soybean functional genomics arena, in addition to the conventional assays involving qRT-PCR, Northern blotting, global gene expression analysis like Serial analysis of gene expression (SAGE), microarrays kind strategies are being widely employed. With the identification of micro RNAs (miRNAs) as ultimate gene effector molecules identification and characterization of novel miRNAs in soybean is gaining a momentum. Thus the rapid development of soybean genomics and transcriptomics has provided tremendous opportunity for the genetic improvement of soybean.

Keywords *Glycine max* • Genome mapping • Genome structure • Comparative genomics • Functional genomics • Diversity • Soybean genetics

Introduction

The legumes are critical in global agriculture, providing the majority of plant protein, and more than a quarter of the world's food and animal feed. Soybean (*Glycine max* L. Merr.) is a member of the legumes, the third largest family of flowering plants. It is a good source of both protein (40 %) and oil (20 %). In the international world trade markets, soybean is ranked number one (53 %) among the major oil crops such as rapeseed, groundnut (peanut), cottonseed, sunflower seed, linseed, sesame seeds, and safflower. Soybean was domesticated in northeastern China about 2500 BC and subsequently spread to southern China, Korea, Japan, and other countries in South-Eastern Asia. Soybean is a self-pollinated diploid and has a chromosome number of $2n = 4x = 40$. Taxonomically, soybean is classified in the legume family, Leguminosae, subfamily Papilionoideae, tribe Phaseoleae, and genus *Glycine*.

Soybean Genome and Its Size

The genus *Glycine* contains two subgenera *Soja* and *Glycine*. Sub-genus *Glycine* contains 26 perennial species which are geographically distributed in Australia. Sub-genus *Soja* contains two species *viz. max*, a cultivated annual species and *soja*, a wild annual species (Table 3.1). Genetic diversity in genus *Glycine* is covered in great detail by Ratnaparkhe et al. (2011). Goldblatt (1981) has reported that the base number for Phaseoleae is almost certainly $x = 11$ and aneuploid reduction to $x = 10$ is prevalent throughout the Papilionoideae including genus *Glycine*. However, Darlington and Wylie (1955) proposed an $x = 10$ basic chromosome number for the cultivated soybean. Based upon these views and on recent taxonomic, cytological

Table 3.1 List of subgenus, species, 2n number, genome symbol and distribution of genus *Glycine*

S. no.	Species	2n	Genome	Geographic distribution
	Subgenus *Glycine*			
1.	*G. albicans* Tind. And Craven	40	I	Australia
2.	*G. aphyonota* B. Pfeil	40	I_3	Australia
3.	*G. arnarea* Tind.	40	H	Australia
4.	*G. argyrea* Tind.	40	A_2	Australia
5.	*G. canescens* F. J. Herman	40	A	Australia
6.	*G. clandestina* Wendl.	40	A_1	Australia
7.	*G. curvata* Tind.	40	C_1	Australia
8.	*G. cyrtoloba* Tind.	40	C	Australia
9.	*G. falcate* Benth.	40	F	Australia
10.	*G. gracei* B. E. Pfeil and Craven	40	?	Australia
11.	*G.hirticaulis* Tind. and Craven	40	H_1	Australia
		80	?	
12.	*G. lactovirens* Tind. and Craven	40	I_1	Australia
13.	*G. latifolia* (Benth.) Newell and Hymowitz	40	B_1	Australia
14.	*G. latrobeana* (Meissn.) Benth.	40	A_3	Australia
15.	*G. microphylla* (Benth.) Tind.	40	B	Australia
16.	*G. montis-douglas* B. E. Pfeil and Craven	40	?	Australia
17.	*G. peratosa* B. Pfeil and Tind.	40	A_5	Australia
18.	*G. pescadrensis* Hayata	80	AB_1	Australia
19.	*G. pindanica* Tind. and Craven	40	H_2	Australia
20.	*G. pullenii* B. Pfeil	40	H_3	Australia
21.	*G. rubiginosa* Tind. and B. Pfeil		A_4	Australia
22.	*G. stenophita* B. Pfeil and Tind.	40	B_3	Australia
23.	*G. syndetika* B. Pfeil and Craven.	40	A_6	Australia
	G. dolichocarpa Tateishi and Ohashi	80	D_1A	Taiwan
24.	*G. tabacina* (Labill.) Benth.	40	B_2	Australia
		80	Complex	
25.	*G. tomentella* Hayata	38	Complex	Australia
		40	Complex	
		78	Complex	
		80		
	Subgenus *Soja*			
26.	*G. soja Sieb. and Zucc.*	40	G	China, Japan, Korea, Russia, Taiwan
27.	*G. max* (L.) Merr.	40	G_1	Cultigen

Source: Ratnaparkhe et al. (2011)

and molecular systematic research on the genus *Glycine* and allied genera, it has been hypothesized that soybean genome is the product of a diploid ancestor ($n = 11$), which underwent aneuploid loss ($n = 10$) and polyploidization ($2n = 40$) (Singh and Hymowitz 1988; Blanc and Wolfe 2004). Based on the chromosomal evidences it has been inferred that polyploidization event was most probably an allopolyploidization event (Gill et al. 2009). Two genome duplications or hybridizations may have occurred (Blanc and Wolfe 2004; Shoemaker et al. 1996, 2002; Tian et al. 2004)

and the duplicated regions might have been segmented and reshuffled (Grant et al. 2000; Yan et al. 2003). Soybean (*Glycine max* L. Merr.) has a genome size of 1.1–1.115 Gb (Arumuganathan and Earle 1991; Schmutz et al. 2010).

The soybean genome is a largest of plant genome, thus far, that underwent sequencing following WGS approach and the outcome match up to all the high quality WGS sequenced plant genomes. The uncovering of draft genome sequence assembly of soybean cultivar Williams 82 following whole-genome shot gun (WGS) approach by Department of Energy-Joint Genome Initiative (DOE-JGI) revealed 950 Mb (megabases) of assembled and anchored sequence as against the predicted 1,115 Mb genome thus representing 85 % of the whole genome (Schmutz et al. 2010). Consequently the size of the genome is significantly larger than the genomes of grapes (505 Mb), *Arabidopsis* (157 Mb), rice (389 Mb), and poplar (485 Mb) and to some extent comparable in size to pigeonpea (833.07 Mb) and tomato (900 Mb) genomes. The soybean genome sequence information has been assembled on to the 397 sequence scaffolds of 20 chromosomes representing the 20 linkage groups of the crop. With the aid of high density genetic maps comprising SNPs (4,991), SSRs (874) among others, the order and orientation of the scaffold placements were accomplished. All of the 397 sequence scaffolds, excluding 20, have been oriented unequivocally on the chromosomes. The assembly features also discloses that the unoriented scaffolds are part of repetitive regions of the genome which are ascribed with low or no recombination frequency and scanty availability of genetic markers.

A striking feature of the soybean genome is that 57 % of the genomic sequence occurs in repeat-rich, low-recombination heterochromatic regions surrounding the centromeres. The average ratio of genetic to-physical distance is 1 cM per 197 kb in euchromatic regions, and 1 cM per 3.5 Mb in heterochromatic regions. These proportions are similar to those in sorghum, in which 62 % of the sequence is heterochromatic, and different than in rice, with 15 % in heterochromatin. Ninety-three percent of the recombination occurs in the repeat-poor, gene-rich euchromatic genomic region that only accounts for 43 % of the genome. Nevertheless, 21.6 % of the high confidence genes are found in the repeat- and transposon-rich regions in the chromosome centres. Schmutz et al. (2010) have identified 46,430 high-confidence protein-coding loci in the soybean genome, using a combination of full-length complementary DNAs, expressed sequence tags, homology and *ab initio* methods. Another, 20,000 loci were predicted with lower confidence; this set is enriched for hypothetical, partial and/or transposon-related sequences, and possess shorter coding sequences and fewer introns than the high-confidence set. Of the 46,430 high-confidence loci, 34,073 (73 %) are clearly orthologous with one or more sequences in other angiosperms, and can be assigned to 12,253 gene families. A combination of structure-based analyses and homology-based comparisons identified 38,581 repetitive elements, covering most types of plant transposable elements (Schmutz et al. 2010). These elements, together with numerous truncated elements and other fragments, make up 59 % of the soybean genome. Long terminal repeat (LTR) retrotransposons are the most abundant class of transposable elements. The intact element sizes range from 1 to 21 kb, with an average size of 8.7 kb. Of the 510 families containing 14,106 intact elements, 69 % are Gypsy-like and the remainder Copia-like.

Comprehensive annotation of the genome disclosed the presence of 32,370 LTR-retrotransposons, 182 LINEs and 6,029 DNA transposons (Du et al. 2010). Besides the list of transposons is endless with the new and unique TE is discovered aftermath of the re-sequencing of the some soybean germplasm lines by Lam et al. (2010). The re-sequencing project also throws light on the soybean population statistics parameter linkage disequilibrium (LD), wherein high LD is detected in both the cultivated and wild genotypes of soybean (Lam et al. 2010). Thus the high LD attribute of soybean genome makes the marker assisted breeding a less challenging breeding method whereas it exerts a limit on QTL mapping or association studies. Nevertheless the newly identified 205,614 tag SNPs (Single nucleotide polymorphisms) may be supportive for these population genetics studies (Lam et al. 2010). The analysis on the soybean genome also revealed that it is exemplified by the presence of large effect single nucleotide polymorphisms (SNPs) and hence the higher non-synonymous to synonymous mutation ratio (Non-syn/Syn) thereby paving way for rapid accumulation of deleterious mutations (Lam et al. 2010).

Genomic Resources

The resources for soybean genomics are aplenty online for the benefit of soybean research community as enumerated in Table 3.2.

Mapping Populations

Various mapping populations in soybean have been developed independently based upon the interests and needs of individual researchers, i.e., the degree of polymorphism required and specific agronomic traits for analysis. F_2 populations or recombinant inbred lines (RILs) have been employed for the construction of linkage maps in soybean. Genetic markers often show polymorphism in one population but not in another population, which hinders the efficient use of the developed markers. While interspecific mapping populations contributed enormously to the saturation of the soybean linkage map, intraspecific linkage maps have also been developed.

Molecular Markers

In order to perform genetic and genomic analysis of the soybean genome, various types of molecular markers have been developed and utilized. The first report of utilization of molecular markers in soybean began with the application of restriction fragment length polymorphism (RFLP) for the assessment of molecular genetic diversity of the soybean nuclear genome (Apuya et al. 1988). Subsequent marker

Table 3.2 Genomic resources available online for soybean improvement

S. no.	Genomic resource	Features/data sets	Bioinformatics tools	Utility	Reference
1.	Phytozome http://www.phytozome.net/ Joint project of the Department of Energy's Joint Genome Institute and the Center for	Genome sequence Gene annotation	Gbrowse BLAST BLAT	Genes/gene families and their evolutionary history Manipulation and display of genome annotation Comparative genomic studies Biomart-data retrieval tool	Goodstein et al. (2012)
2.	Soybase and soybean breeder's tool box http://soybase.org/ USDA and Iowa State University	Genetic and physical maps Genome sequence and annotation Transposable elements TE (data base)	BLAST Search EST library SoyChip annotation, Probe analysis pHap-potential haplotype search Metabolic pathway database Transcriptome of tissue based expression RNA-Seq Atlas Mutants browsers	A comprehensive repository provides curated data on genetics, genomics and related area thereby integrating soybean genetics and molecular biology	Grant et al. (2010), Du et al. (2010)
3.	Soy knowledge base http://soykb.org/ University of Missouri, Columbia	Genes, gene families, homologous genes SNPs miRNAs 3-D protein structure Metabolites	BLAST Gene pathway viewer Metabolic pathway viewer Affymetrix probeID Mapper Motifsampler by WebLOGO Chromosome Visualizer tool	SoyKB provides data integration from gene through phenotype encompassing entire soybean Omics thus a web resource for soybean translational genomics.	Joshi et al. (2012)

4.	Soy-TFKB (Soybean Transcription Factor Knowledge Base) http://www.igece.org/Soybean_TF/ Institute for Green Energy and Clean Environment	Transcription factors Transcription factors (TF) families	TF browser	Complete Transcription Factor Data base (TFDB) which is genome based as against other EST or scaffold based TF classifications	http://www.igece.org/Soybean_TF/
5.	Soy DB http://casp.rnet.missouri.edu/soydb/ National Science Foundation and University of Missouri	Amino acid sequences Predicted tertiary structures DNA binding sites and motifs Domain predictions Homologous proteins from the Protein Data Bank Protein family classification	Position-Specific Iterated BLAST (PSI-BLAST) Protein family search Multiple sequence alignments Protein annotation into known 64 TF families by Hidden Markov Model (HMM)	A user friendly database containing protein annotation based comprehensive knowledge on soybean transcription factors.	Wang et al. (2010)
6.	SGMD (The soybean genomics and microarray database) http://bioinformatics.towson.edu/SGMD/ Housed in the Matthew's laboratory of Soybean Genomics and Improvement Laboratory (SGIL) in collaboration with Beltsville Agricultural Research Center (BARC) and others	Genomic EST and Microarray data	Local Blast Search Accelerator (LOBSA)	The resource is meant for researchers, interested in soybean and soybean-pest/pathogen interactions, to collaborate. Microarray experiment details in soybean on SCN infection	Alkharouf and Matthews (2004)

(continued)

Table 3.2 (continued)

S. no.	Genomic resource	Features/data sets	Bioinformatics tools	Utility	Reference
7.	Soybean genome map http://soybeangenome.siu.edu/ Southern Illinois University Carbondale	*G. max* (SIU 2005) version 4	Key word search Clones, Marker loci, contigs, ESTs, sequence related to soybean genetic map	Gbrowser enabled search for the soybean map relevant informations	–
8.	miRBase http://www.mirbase.org/ Faculty of Life Sciences, University of Manchester	miRNA (hairpin sequence and mature miRNA) sequences miRNA targets (micro-Cosm database)	miRNA search by name, key word, accession, tissue specific expression, for clusters, by sequence information Downloadable miRNA sequence information in EMBL, FASTA format, family classification of miRNAs	The resource aids in computational validation of miRNAs, secondary structure characterization, target transcripts prediction and thus maintenance of repository of small RNAs	Kozomara and Griffiths-Jones (2011)

analysis employed RFLP, random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs). Apuya et al. (1988) analyzed randomly chosen 300 RFLP probes in genomic DNA of the genetically distant cultivars Minosy and Noir 1. One in five probes revealed a polymorphism, more than half of these polymorphisms appear to result from rearrangement of the genomic DNA. Twenty seven markers were analyzed for linkage in F_2 plants and eleven of these markers were present in four linkage groups. Subsequently RFLP markers were used extensively for genetic diversity analysis (Keim et al. 1989, 1992; Skorupska et al. 1993; Lorenzen et al. 1995) and linkage mapping (Keim et al. 1990, 1997; Diers et al. 1992a; Lark et al. 1993; Akkaya et al. 1995; Shoemaker and Specht 1995; Mansur et al. 1996; Cregan et al. 1999; Ferreira et al. 2000; Yamanaka et al. 2001; Lightfoot et al. 2005) until SSR and SNP markers have become popular. Owing to simple method of detection and no prior knowledge of DNA sequence requirement, RAPDs were also used extensively by soybean geneticists, mainly for germplasm classification (Thompson et al. 1998; Brown-Guedira et al. 2000; Li and Nelson 2002). A large number of AFLP markers were also utilized for linkage map construction in soybean (Keim et al. 1997; Matthews et al. 2001).

The discovery and use of SSR markers in human (Litt and Luty 1989; Tautz 1989; Weber and May 1989) and other mammalian systems (Dietrich et al. 1994) prompted the soybean scientists to use SSRs in soybean. Indeed, the first report of SSR allelic variation and their use as marker system in plant species appeared from soybean (Akkaya et al. 1992; Morgante and Olivieri 1993). The early studies on SSR polymorphism revealed very high level of allelic variation in cultivated and wild soybean genotypes (Maughan et al. 1995; Morgante et al. 1994; Rongwen et al. 1995). Akkaya et al. (1995) first time developed 40 SSRs and integrated them to soybean linkage map. Subsequently, Cregan et al. (1999) developed a large set of SSRs and mapped 606 SSR loci in one or more of the three mapping populations to develop an integrated linkage map. Song et al. (2004) developed 420 new SSRs from ESTs, BAC-end sequences and genomic libraries, and added them to the soybean integrated linkage map developed by Cregan et al. (1999). Hisano et al. (2007) developed SSR markers using publically available EST sequence information. A total of 6,920 primer pairs were designed to amplify SSRs from 63,676 non redundant soybean ESTs. As a result, 668 EST-derived marker loci were mapped on soybean linkage map. With the advancement of sequencing technologies and availability of more and more sequence information, development and use of SSRs increased greatly in soybean. The availability of BAC-end sequence facilitated development of comprehensive sets of SSRs leading to integration of physical map with genetic map (Shultz et al. 2007; Shoemaker et al. 2008). Recently, utilizing the whole genome sequence, a soybean SSR database (BARCSOYSSR_1.0) containing genome position and primer sequences for 33,065 SSRs was developed by Song et al. (2010). This genome wide SSR database is effective in providing informative SSRs at any genomic position required for fine mapping to identify the position of a causative gene as well as for marker assisted selection.

SNPs provide an abundant source of DNA polymorphism in comparison to SSRs, thereby improving the success rate in a diversity of applications including

QTL mapping, positional cloning, association analysis and determination of genetic relatedness among individuals. To determine SNP frequency in coding and noncoding regions of soybean genome, approximately 28.7 kbp of coding sequence, 37.9 kbp of noncoding perigenic DNA, and 9.7 kbp of random noncoding genomic DNA was sequenced in 25 diverse soybean genotypes by Zhu et al. (2003). The nucleotide diversity (θ) observed was 0.00053 and 0.00111 in coding and in nocoding perigenic DNA, respectively, whereas combined nucleotide diversity of whole sequence analyzed was 0.00097. Squared allele frequency correlations (r^2) among haplotypes at 54 loci with two or more SNPs indicated low genome wide linkage disequilibrium. Choi et al. (2007) identified SNPs via the resequencing of sequence-tagged sites (STSs) developed from EST sequences. From an initial set of 9,459 STS primers designed to a diverse set of genes, 4,240 STSs were amplified and sequenced in six diverse soybean genotypes. In the total 2.44 Mbp of aligned sequence, a total of 5,551 SNPs were discovered, including 4,712 single-base changes and 839 indels resulting in an average nucleotide diversity of $\theta = 0.000997$. Exploiting these SNPs a total of 1,141 genes were placed on the genetic map by virtue of a SNP segregating among one or more RIL mapping populations, thus constructed a transcript map in soybean.

Hyten et al. (2008) developed a multiplex assay of 384 SNPs designated as soybean oligo pool all-1 (SoyOPA-1), for genotyping in the complex genome of soybean. This custom 384-SNP GoldenGate assay was designed using SNPs discovered through resequencing of five diverse accessions. SNP allelic data were obtained for 342 of the 384 SNPs in SoyOPA-1 with success rate of 89 %. In order to develop more SNPs Hyten et al. (2010a) sequenced a total of 3,268 SNP-containing robust STS in six diverse genotypes, resulting in identification of 13,042 SNPs with an average of 3.5 SNP per polymorphic STS. These SNPs along with 5,551 SNPs discovered by Choi et al. (2007) were used to design two Illumina custom 1,536 SNP GoldenGate assays designated as SoyOPA-2 and SoyOPA-3. A set of 1,536 SNPs from the 3,456 SNPs present in three SoyOPAs was selected to include sufficient polymorphic SNP markers distributed throughout the genome that could be used for most of the QTL mapping applications. This set of 1,536 SNPs GoldenGate assay was designated as Universal Soy Linkage Panel 1.0 (USLP1.0). Hyten et al. (2010b) sequenced a reduced representation library of soybean to identify SNPs using high throughput sequencing methods. A total of 7,108–25,047 SNPs were detected by using multiple SNP detection methods. A total of 1,536 SNPs were selected from this pool of 7,108 SNPs to create an Illumina GoldenGate assay (SoyOPA-4). The SoyOPA-4 produced 1,254 successful GoldenGate assays indicating a validation and assay conversion rate of 81.6 % for the predicted SNPs. Chaisan et al. (2010) used 335,857 publically available ESTs derived from 18 genotypes for EST clustering and *In silico* SNP identification. A total of 3,219 EST contigs were established based on three to nine genotypes and a total of 26,735 SNPs were identified. The confirmation of *In silico* identified SNPs by Sanger sequencing yielded 15.7 % accuracy rate between two cultivars Williams 82 and Harosoy. These studies resulted in development of large number of SNP markers in soybean which could be utilized for mapping of complex traits as well as molecular breeding applications.

ESTs and Functional Markers

A simple and efficient method for identification of the expressed genes in an organism is random sequencing of gene transcripts also called as expressed sequence tags (ESTs). ESTs are crucial resource for genome annotation, and provide useful information about gene structure, alternative splicing, expression patterns and transcript abundance (Umezawa et al. 2008). In soybean, even before the start of the whole genome sequencing project, large-scale EST-sequencing projects were undertaken, and a large number of ESTs have become available. Presently, the soybean research community possesses a collection of more than 14,00,000 EST sequences (http://www.ncbi.nlm.nih.gov/dbEST/). These ESTs were obtained from cDNA libraries prepared from tissues representing a wide range of plant developmental stages, organs, genotypes, and biotic and abiotic challenges (Shoemaker et al. 2002; Umezawa et al. 2008; Sha et al. 2012). This EST collection represents a large resource of publicly available genic sequences and provides valuable insight into structure, function and evolution of this model crop legume (Shoemaker et al. 2003). EST resources have been also used in applied aspects like functional genomics studies by developing cDNA arrays, and by exploiting them in the development of molecular markers. Using automated procedures, ESTs and full-length mRNA sequences from characterized genes are partitioned into sets, or "clusters," that are very likely to represent distinct genes. Such cluster sequences are called as "unigenes" which are used to select reagents for large-scale gene mapping and gene expression studies (Schuler 1997). The recent clustering of ESTs in NCBI database (UniGene Gma build 42) contain 35,982 unigenes developed by clustering of 13,54,263 ESTs (http://www.ncbi.nlm.nih.gov/UniGene/UGOrg.cgi?TAXID=3847).

Among the popular molecular markers that can be developed from ESTs or unigenes are simple sequence repeats (SSRs), single-nucleotide polymorphisms (SNPs) and conserved orthologous sets of markers (COSs). Putative functions can be deduced for the markers derived from ESTs or genes using homology searches with existing databases. Therefore, molecular markers generated from gene sequence data are known as "functional markers" (FMs) (Varshney et al. 2006). Excellent examples of development of functional SNP markers were reported in two vegetable type soybean cultivars for fragrance alleles of *GmBADH2* gene (Juwattanasomran et al. 2011, 2012). Lipoxygenases are the enzymes responsible for the development of unpleasant flavors in foods containing soybean by oxidation of polyunsaturated fatty acids. Soybean seeds contain three lipoxygenase (Lox) enzymes that are controlled by separate genes, *Lox1*, *Lox2* and *Lox3*. Genetic studies demonstrated that the absence of each enzyme is under the control of three null alleles, *lox1*, *lox2* and *lox3*, which are inherited as simple recessive alleles. Perfect molecular marker assays were designed to distinguish mutant from wild type alleles for *Lox1*, *Lox2* and *Lox3* genes which showed a complete association between the inheritance of homozygous *lox* mutant alleles and the lack of lipoxygenase activity (Lenis et al. 2010). Gillman et al. (2009) have identified recessive mutations in two soybean homologs of the maize *lpa1* gene in soybean line CX1834, a mutagenized

line with a low phytic acid phenotype and consequently developed functional marker assays for *lpa1* and *lpa2* genes that can be used to directly select for the mutant alleles that control the phenotype.

Reducing the saturated fatty acid concentration in soybean oil is required to reduce the risk of coronary heart disease. The primary focus of modifying the saturated fatty acid profile has been on reducing the palmitate concentration. The mutant alleles associated with a decrease in palmitate concentration in soybean seed oil result from mutations in the fatty acid thioesterase B (FAT B) enzyme; specifically the 16:0-ACP thioesterase enzyme (16:0-ACP TE). DNA sequence analysis of *GmFATB1a*, the major 16:0-ACP TE gene of soybean, revealed a single nucleotide polymorphism (SNP) resulting in single amino acid substitution that is likely to be detrimental to enzyme function. A functional marker based on SimpleProbe molecular marker assay was developed for SNP detection in *GmFATB1a* gene, which will be useful for selection of low palmitate genotypes (De Vries 2011).

BAC Libraries

In soybean, several BAC libraries have been developed from different genotypes. These soybean BAC libraries have been developed with different objectives including general genomic research as well as specifically for cloning of disease and insect resistance loci. Various important characteristics and purposes of the selected soybean BAC libraries are summarized in Table 3.3. These libraries will provide a good resource for positional cloning of agronomical and biologically important QTL genes that the representative genotype possesses. Moreover, the ability to introduce large insert DNA clones into plant cells (Hamilton et al. 1996) provided new avenues for functional genomics, genetic engineering of complex loci and the assembly of several unlinked genes into a single locus. Therefore, BAC vectors are designed for cloning large fragments that might contain a gene cluster or an intact locus which may be useful for plant transformation and functional analysis (Meksem et al. 2000a, b). BAC libraries constructed from various cultivars with different backgrounds are very useful, because not all genes necessarily exist in one germplasm (Xia et al. 2005). For example, the resistance genes, *rhg1* and *Rhg4*, to the soybean cyst nematode were isolated by positional cloning from a Forrest BAC library. Zhu et al. (2009) constructed a unique soybean BAC library derived from a soybean genotype PI 229358 that carries defoliating insect-resistance alleles. This library should be useful for the map-based cloning of the major insect resistance quantitative trait loci (QTL), QTL-M, and other insect-resistance QTLs from as well. BAC libraries have also been constructed for *G. soja, G. syndetika, G. canescens, G. stenophita, G. cyrtoloba, G. tomentella, G. falcata*, and the polyploid, *G. dolichocarpa*. All libraries are publicly available through the Arizona Genome Institute and are part of an NSF Plant Genome project to leverage diversity within the genus *Glycine*.

Table 3.3 Characteristic features of soybean BAC libraries

S. no.	Soybean cultivar used	Number of BAC clones	Average insert size (kbp)	Purpose of BAC library construction	References
1	Williams 82	40,000	150	Genomic research	Marek and Shoemaker (1997)
2	Faribault	30,720	120	Cloning SCN[a] resistance gene	Danesh et al. (1998)
3	Williams 82	45,000	105	Cloning*Rps1*-k gene	Salimath and Bhattacharyya (1999)
4	PI 437654	73,728	136	Cloning SCN resistance gene	Tomkins et al. (1999)
5	A3244	206,592	148	Genomic research	Tomkins et al. (2000)
6	Forrest (two libraries)	38,400	125	Genomic research, SCN and SDS[b] resistance	Meksem et al. (2000b)
7	Forrest	38,400	157	Genomic research	Wu et al. (2004b)
8	Misuzudaizu	53,760	116	Genomic research	Xia et al. (2005)
9	PI 229358	55,296	131	Insect-resistance QTL	Zhu et al. (2009)

[a]*SCN* soybean cyst nematode
[b]*SDS* sudden death syndrome

Physical Mapping

Availability of genomic clone libraries with large DNA inserts is one of the essential requirements for plant genome analysis, primarily for physical mapping, gene isolation, and gene structure and function analysis. Various cloning vectors have been extensively used in generating genomic libraries with large DNA inserts. The bacterial artificial chromosome (BAC) vectors have been used widely for generating genomic DNA libraries in economically important crop plants including soybean. Development of BAC libraries is considered as critical step towards physical mapping and positional cloning of important genes.

Physical Map

The genome wide physical maps are an important platform for the genomic researches in an economically important crop like soybean. Besides assisting in the whole genome sequencing venture it facilitates the researchers in different applications like targeted genetic marker development, accelerates positional cloning approaches and rapid and robust EST mapping as well. The physical map generation of soybean was initiated with the development of early genetic maps characterized by the even distribution on the whole genome of the crop. Yeast artificial chromosomes (YAC) were initially developed with a view to utilize the resource for chromosome walking and *in situ* hybridization (Zhu et al. 1996). Bacterial artificial chromosome (BAC)

libraries covering the whole soybean genome were generated by early genomic researchers (Marek and Shoemaker 1997; Danesh et al. 1998; Tomkins et al. 1999; Salimath and Bhattacharyya 1999; Meksem et al. 2000a). Notwithstanding the small size of the predicted soybean genome, when compared with other organisms for which whole genome physical maps have been developed, the endeavor was an overwhelming task considering the fact that the genome underwent large scale duplications twice ~59 and 13 million years ago (mya), respectively (Shoemaker et al. 2006; Schmutz et al. 2010). In consequence the soybean is a paleopolyploid genome, typified by duplication and diploidisation, resulting in highly replicated genome and thus has on average of 2.55 duplicated segments (Shoemaker et al. 1996). Nevertheless, BAC libraries encompassing variety of genotypes in combination with diverse enzymes, have led to the development of early physical contigs (Marek and Shoemaker 1997). Efforts were made to develop physical map of soybean genome on a regional level using BAC based libraries but the map could possibly cover only 20 % of the genome (Meksem et al. 2001). It was followed by the construction of genome-wide physical map utilizing 78,001 BAC and Binary large-insert BAC clones (BIBAC clones) representing 9.6 haploid genomes and three cultivars of soybean (Wu et al. 2004a). The map comprises 2,905 BAC/BIBAC contigs, estimated to embody 1,408 Mb in physical length thus pointing towards the overlapping nature of the contigs employed in the generation of the map. Greater proportion of the physical map was anchored to the genetic map with the support of RFLP and SSR markers available by then (Wu et al. 2004a). A physical map of soybean cultivar Williams 82 was in place that was generated from 67,968 BAC clones from a *Bst*Y I library and 40,320 clones from a *Hin*d III library (http://soybeanphysicalmap.org). The physical map was developed using high information content fingerprint (HICF) approach (Luo et al. 2003; Warren and The Soybean Mapping Consortium 2006) with the BAC clones assembled in to 1,893 contigs and around 3,000 singletons (http://www.soybase.org). Furthermore in a quest to develop well developed physical map, overlaid on to a sequence based genetic map, SSR markers derived from BAC ends sequence (BES) were mapped and integrated in to the physical map to improve its quality (Shoemaker et al. 2008). Despite the complexity of the soybean genome six dimensional BAC clones pools were employed to demonstrate the anchoring of genetic markers to the soybean BAC clones (Wu et al. 2008). The 6-D pool screening endeavor resulted in anchoring of 1,470 markers (580 SSRs and 890 STSs) on a *Bst* Y I BAC library generated from cultivar Williams 82. The physical framework comprises more than 7,000 BAC clones, anchored employing 1,470 markers, representing the complete genome (Wu et al. 2008). On the parallel lines soybean unigene sets from NCBI were computationally anchored to Williams 82 BAC end sequences (BES) resulting in anchoring of additional 305 contigs thereby complementing 1,184 anchored contigs achieved through 6-D pool screening efforts (Wu et al. 2008). Thus the physical framework was accomplished by associating the contigs to the molecular markers which in turn was achieved by finger printing of the BAC clones through overgo hybridization, RFLP hybridization and SSR amplification (Song et al. 2004; Choi et al. 2007). Thus the soybean physical map has been updated to the final version (Oct 2008) with the contribution from 86,524 soybean

BACs as well as HICF fingerprints from 37,658 BACs (http://soybeanphysicalmap.org). The soybean physical map is made available public under Soybean Breeders Toolbox (SBT) set up in soybase website (http://www.soybase.org) for the greater benefit of research community.

Whole Genome Sequencing and Data Mining

Soybean genome sequencing project was accomplished by US Department Of Energy-Joint Genome Initiative (DOE-JGI)-Community Sequencing Program (CSP). The endeavor was carried out by traditional Sanger's method by whole genome shotgun sequencing approach where in the entire genome is randomly sheared, subcloned and sequenced redundantly. The strategy is a preferred method of choice owing to its relative ease, cost-effectiveness and rapidity. The present genome sequence assembly made available public is a first chromosome scale assembly and termed as Glyma-1.0. This reveals approximately 950 Mb genome of expected 1,115 Mb assembled on to 20 chromosomes of the crop. The chromosome scale assembly was generated by employing assembler Arachnae 2, suitable for repetitive genomes like soybean, at JGI-Stanford Human Genome Centre. Then integration of sequence information with the genetic and physical maps already available to obtain glyma1 reads were carried out at the University of Minnesota (http://www.phytozome.net/soybean). The protein coding regions have been predicted to be 66,153, of which over 46,000 genes are predicted with high confidence level (Schmutz et al. 2010). Thus with reference to the "gene space" the assembly is complete in all aspect as comparisons with EST database on soybean revealed the 98 % coverage of the protein coding genes. In terms of accuracy the Glyma 1 is highly accurate in the genic regions as EST sequences matches exactly with the assembled genome sequences. Moreover any discrepancies that arise between the shot-gun assembly and the studies based on physical or genetic maps have been eradicated manually to make the Glyma 1 error free with respect to large scale genome structure of soybean (Schmutz et al. 2010).

Glyma.1 gene set was prepared employing homology based computational prediction algorithms like GenomeScan from Chris Burge and FgenesH predictions provided by Asaf Salamov at JGI, along with the PASA program to integrate soybean ESTs. Peptides from other flowering plants, TIGR legume EST data base were used and aligned with soybean genome data to obtain the gene rich regions. The resultant regions were fed in to the gene prediction algorithms to find putative genic regions. The homologous regions were integrated with EST sequences using PASA program (Haas et al. 2003).

The genome sequence data and gene annotation of soybean, among other 30 green plants, is housed in Phytozome v8.0 database (http://www.phytozome.net/) which is a joint project of the Department of Energy's Joint Genome Institute and the Center for Integrative Genomics along with the University of California. It provides access to genes and gene families either by keyword based search or sequence similarity based programs like BLAST and BLAT (BLAST like Alignment Tool).

The sequence analysis via shared functional domain or consensus sequence similarity enables the study on the evolutionary history of each gene family and identification of the closely linked gene families. Gbrowser in the database facilitates EST alignments, utility of VISTA tracks that helps in assessing the extent of nucleotide conservation in related plant genera. The Biomart-open source data retrieval software allows the research community to download complete data from phytozome.

Genetic Mapping

An accurate and well-saturated genetic linkage map is fundamental to modern plant genetics and breeding. A genetic map allows the identification of genomic loci controlling an agronomic trait, including quantitative trait loci, and an understanding of genetic diversity and genome structure of genetic resources. Furthermore, such a linkage map is required for anchoring of a physical map. The first report of construction of genetic linkage map in soybean using molecular markers was published by Keim et al. (1990) (Table 3.4). Using a F_2 population of 59 plants derived from a cross between A81-356022 × *Glycine soja* (PI468916) (A × PI), a total of 150 RFLP markers were mapped on soybean genome. The map contains 26 linkage groups covering 1,200 cM length. Subsequently, a large number of RFLP markers were added to soybean linkage map by Diers et al. (1992a), Lark et al. (1993), Akkaya et al. (1995), Shoemaker and Specht (1995), Mansur et al. (1996), Cregan et al. (1999), Ferreira et al. (2000), and Yamanaka et al. (2001) (Table 3.4). Later on, RAPD and AFLP markers were also utilized for genetic mapping in soybean (Ferreira et al. 2000; Keim et al. 1997; Matthews et al. 2001) (Table 3.4) until SSRs and SNPs become popular and marker of choice. Keim et al. (1997) mapped a total of 650 AFLP loci in the 42 RILs of the cross PI437654 × BSR101. Ferreira et al. (2000) mapped 106 RAPD markers in the same mapping population along with 250 RFLP markers. Matthews et al. (2001) mapped a total of 105 AFLP markers in a F_2 population of 149 plants derived from cross Noir 1 × BARC-2.

The high level of polymorphism combined with the random distribution in genome as well as their single locus nature and simple method of detection suggested that SSRs were excellent complement to RFLP markers for use in soybean genetics, genomics and breeding research (Cregan 2008). Akkaya et al. (1995) first time mapped 34 SSRs and integrated them to soybean RFLP linkage map of an aF_2 population derived from cross Clark × Harosoy (Table 3.4). Subsequently, Cregan et al. (1999) mapped 606 SSR loci in one or more of the three mapping populations to develop an integrated linkage map (Table 3.4). These three different mapping populations were: the USDA/Iowa State *G. max* × *G. soja* F_2 population, the Univ. of Utah Minsoy × Noir1 (M × N) recombinant inbred population, and the Univ. of Nebraska Clark × Harosoy F_2 population. Each SSR loci mapped to a single genomic location with map order essentially identical in all three populations. This integrated linkage

Table 3.4 A summary of molecular linkage maps developed in soybean

Mapping population used	Population size and type	No. of loci	Types of DNA marker used (No.)					Genetic map length (cM)	References
			RFLP	RAPD	AFLP	SSR	SNP		
A81-356022 × PI468916	59 F_2 plants	150	150					1,200	Keim et al. (1990)
	59 F_2 plants	238	238					2,147	Diers et al. (1992b)
	59 F_2 plants	368	358	10				2,473	Shoemaker (1995)
	59 F_2 plants	1,008	501	10	11	486		3,003	Cregan et al. (1999)
Minsoy × Noir 1	69 F_3 families	132	132					1,551	Lark et al. (1993)
	240 RILs	265	224			41		1,981	Mansur et al. (1996)
	240 RILs	621	209			412		2,413	Cregan et al. (1999)
Clark × Harosoy isolines	57 F_2 plants	83	80	3				1,004	Shoemaker (1995)
	57 F_2 plants	118	80	4		34		1,486	Akkaya et al. (1995)
	57 F_2 plants	514	95	69	11	339		2,787	Cregan et al. (1999)
PI437654 × BSR301	42 RILs	840	165	25	650			3,441	Keim et al. (1997)
	42 RILs	356	250	106				3,275	Ferreira et al. (2000)
Changnong 4 × Xinmin 6	88 RILs	237	100	62	42	33		3,713	Liu et al. (2000)
Misuzudaizu × Moshidou Gong 503	190 F_2 plants	498	401	1		96		2,909	Yamanaka et al. (2001)
	94 RILs	935	105			829		2,700	Hisano et al. (2007)
Noir 1 × BARC-2	149 F_2 plants	186	39	17	105	25		1,400	Matthews et al. (2001)
Essex × Forrest	100 RILs	337	41	90		206		2,823	Lightfoot et al. (2005)
Williams 82 × PI 468916	444 RILs	1,790					1,790	2,537	Hyten et al. (2010b)
Consensus maps	Multiple populations	1,803	709	73	6	1,015		2,524	Song et al. (2004)
	Multiple populations	2,982	634	4		983	1,361	2,389	Choi et al. (2007)
	Multiple populations	1,810				1,810		2,443	Hwang et al. (2009)
	Multiple populations	5,500	664			1,006	3,792	2,296	Hyten et al. (2010a)

map consists a total of 1,423 marker loci including 606 SSRs, 689 RFLPs and 26 classical loci aligned into 20 sets of linkage groups. Song et al. (2004) mapped 420 new SSRs and added them to the soybean integrated linkage map developed by Cregan et al. (1999) to construct a new integrated genetic linkage map (Table 3.4). Using one or more of the three mapping populations used by Cregan et al. (1999) as well as two additional RIL populations from Univ. of Utha; Minsoy × Archer (M × A) and Archer × Noir 1 (A × N), a consensus map was developed using JoinMap software. This new integrated genetic map covers 2,523.6 cM of map distance across 20 linkage groups that contained 1,849 markers, including 1,015 SSRs, 709 RFLPs, 73 RAPDs, 24 classical traits, six AFLPs and ten isozymes (Table 3.4). Later on, additional SSR markers were mapped on soybean genome by Shultz et al. (2007) and Shoemaker et al. (2007), to fill the large gaps and to anchor genetic maps with physical map (Table 3.4).

Expressed sequence tag (EST)-derived SSRs are particularly popular because they are cost effective and easy to develop, informative, display putative functional polymorphism and transferable to other species for comparative mapping. In soybean, Hisano et al. (2007) generated a high-density genetic linkage map of soybean using EST-derived SSR markers (Table 3.4). A total of 6,920 SSR primer pairs were designed from 63,676 publicly available non-redundant soybean ESTs. Primer pairs showing polymorphism were then used for genotyping 94 RILs derived from a cross between the Japanese cultivar "Misuzudaizu" and the Chinese line "Moshidou Gong 503" (M × M). A total of 693 polymorphic SSR loci were detected using the 668 EST-derived microsatellite markers, which were used along with 242 other marker loci to develop a high density genetic map spanning 2,700.3 cM of map length (Table 3.4). Hwang et al. (2009) mapped a total of 1,810 SSRs including 693 EST-derived SSR loci, in one or more of three recombinant inbred populations; the US cultivar "Jack" × the Japanese cultivar "Fukuyutaka," the Chinese cultivar "Peking" × the Japanese cultivar "Akita," and the M × M population used by Hisano et al. (2007) (Table 3.4). The integrated linkage map span 2,442.9 cM of genetic map length with the average number of molecular markers per LGs was 90.5 (range of 70–114). The presence of SSRs in EST sequence provides one means for the genetic mapping; however, the number of polymorphic SSRs present in ESTs appears rather limited.

Alternatively, discovery of SNPs in genic sequence would provide a good source of markers, because SNPs are more abundant than SSRs, they improve the odds of success in a diversity of applications including genetic mapping, positional cloning, association analysis, and QTL mapping. In order to construct a transcript map of soybean, Choi et al. (2007) developed STSs primers using ESTs and 3′-unigene sequence and utilized these STSs for the discovery of SNPs via the resequencing of six diverse soybean genotypes. A total of 5,551 SNPs were discovered and 1,141 genes were placed on the genetic map by virtue of a segregating SNP mapped in one or more of the three RIL mapping populations: the University of Utah M × N, and M × A as well as the Evans × PI 209332 (E × PI) (Table 3.4). The analysis of the theoretical distribution of map distances between adjacent genic sequences within

linkage groups clearly indicated clustering of genes. This transcript map of 1,141 genes was integrated with genetic map of 1,015 SSRs developed by Song et al. (2004). Hyten et al. (2008) further added 334 SNPs to this map using 384 SNP GoldenGate genotyping assay (SoyOPA-1). This new integrated consensus map should enhance both applied and basic soybean genetics and genomics research, including QTL discovery, marker-assisted selection, map-based cloning, and the anchoring of the physical to the genetic map.

Hyten et al. (2010a) used three Illumina GoldenGate assays namely SoyOPA-1, SoyOPA-2 and SoyOPA-3 consisting of 3,456 SNP loci for genetic mapping in three mapping populations: the Univ. of Utah "M×N," "M×A" and Univ. of Minnesota "Evans"×"Peking" (ExP). In total, the three SoyOPAs included 2,651 new SNP markers which were segregating in one or more of the three different mapping populations. A fourth version of the soybean integrated genetic linkage map (Consensus Map 4.0) was created by combining the SNP locus data of the E×PI and A×PI mapping populations used by Choi et al. (2007) with the new SNP locus data of the M×N, M×A, and E×P mapping populations to create an integrated genetic linkage map of 5,500 markers spanning a genetic map distance of 2,296.4 cM (Table 3.4). A set of 1,536 SNPs were selected from the 3,456 SNPs present in the three SoyOPAs to create a "Universal Soy Linkage Panel" (USLP 1.0). SNPs for USLP 1.0 were selected based on even distribution throughout each of the 20 consensus linkage groups and to have a broad range of allele frequencies in diverse germplasm. The 1,536 USLP 1.0 will allow fast genotyping and creation of a comprehensive genetic map in most QTL mapping populations and thus will serve as a useful tool for high-throughput QTL mapping.

Hyten et al. (2010b) reported deep resequencing of a reduced representation library for high-throughput SNP discovery, which were used to create a high resolution map required for anchoring and orienting additional scaffolds in the soybean whole genome sequence. In total, 7,108 SNPs were predicted for use in anchoring and orienting additional scaffolds. Ultimately, 1,536 SNPs were selected from this pool of 7,108 SNPs to create an Illumina GoldenGate soybean oligo pool all (SoyOPA-4). SoyOPA-4 was used to genotype 470 F_5-derived RILs from the Williams 82×PI 468916 (W82×468) population along withSoyOPA-3. A total of 550 polymorphic SNPs from SoyOPA-3 and 1,240 polymorphic SNPs from SoyOPA-4 were mapped using 444 RILs to create the 20 linkage groups with an estimated total genetic length of 2,537 cM. The high-resolution W82×468 genetic map containing 1,790 SNP markers was successful in anchoring and orienting additional scaffolds in the 8× scaffold assembly. It added new markers to 335 8× assembly scaffolds, of which 23 scaffolds were previously unmapped in the preliminary 6.5× scaffold assembly. To conclude, several genetic maps have been developed for soybean with various types of molecular markers with recent trend towards development of high density SSR and SNP based maps. These well saturated or high-density linkage maps should facilitate ongoing and future genomic research such as quantitative trait loci mapping and positional cloning in addition to marker-assisted selection in soybean breeding.

QTLs

Quantitative Trait Loci (QTL) Mapping

Different type of molecular markers has been used to map genomic location of major genes and quantitative trait loci (QTLs) for many traits of agronomic and economic importance in soybean. More than thousand QTLs representing more than 90 agronomically important traits have been mapped in soybean (Grant et al. 2010). Current information on all mapped QTLs in soybean is available on the USDA-ARS soybean genetic database *SoyBase* (http://soybase.org). Although a number of QTLs were mapped in the soybean but introgression and pyramiding of genes or QTLs affecting the same trait is a great challenge to breeding programs. For molecular breeding applications, the QTL data published will be useful only if QTL can be validated in independent mapping population(s). Fortunately, many of the QTLs were validated in different genetic backgrounds or in the independent studies and given the name confirmed QTLs "cq" by Soybean Genetics Committee. An attempt has been made to summarize the confirmed QTLs by a review of available information in *SoyBase* and published literature which is presented in Table 3.5.

A total of 25 QTLs were identified to be validated in different genetic backgrounds and thus they can be utilized confidently in marker assisted breeding. The highest percentage of QTLs confirmed was for pest resistance (44 %), probably due to the high heritability of pest resistance. Soybean aphid (*Aphis glycines* Matsumura) is a major sucking pest in soybean and heavy aphid infestation could cause considerable yield loss, especially when aphid density peaks at the beginning of flowering. Three QTLs on three different linkage groups have been validated for aphid resistance in soybean. Zhang et al. (2009) validated two aphid resistance QTLs identified on linkage groups F and M. A mapping population of 51 F_3-derived lines developed from the cross PI 567541B (resistant) × E00003 (susceptible), along with 50 advanced breeding lines was used for the QTL validation. QTL analysis identified two QTLs, detected at similar genomic regions as in the original mapping population. The two QTLs combined with their interaction explained 95.2 % of the phenotypic variation in the field trial. Another aphid resistance locus on linkage group J has been validated by Zhang et al. (2010) in a population of 96 $F_{4:5}$ lines from a cross between PI 567543C and "Skylla," where Skylla is an aphid-susceptible cultivar. The QTL was detected at similar position in the validation population and explained the majority of the phenotypic variation in the field trial.

Soybean cyst nematode (SCN, *Heterodera glycines* Ichinohe) which is the most destructive pest of soybean worldwide, studied extensively for QTL mapping. The QTLs for SCN resistance have been validated for seven genomic locations (Table 3.5). The sudden death syndrome (SDS) of soybean is caused by *Fusarium solani*. The use of resistant cultivars is the most effective method for controlling SDS, therefore mapping SDS resistance QTL is of prime importance. One QTL was mapped and confirmed for SDS resistance by Farias Neto et al. (2007). This QTL on linkage group D2 was tested in a population of F_2 plants developed through one

Table 3.5 List of selected QTLs reported in soybean

Traits	Populations (No.)	Linkage group	No. of QTLs confirmed	References
Soybean cyst nematode resistance	6	A2	1	Matthews et al. (1998); Prabhu et al. (1999); Meksem et al. (2001)
	1	E	1	Kabelka et al. (2005)
	8	G	3	Meksem et al. (2001); Glover et al. (2004); Kabelka et al. (2005); Vuong et al. (2010)
	3	J	1	Glover et al. (2004)
	1	O	1	Vuong et al. (2010)
Soybean aphid resistance	1	J	1	Zhang et al. (2010)
	1	F	1	Zhang et al. (2009)
	1	M	1	Zhang et al. (2009)
Sudden death syndrome	2	D2	1	Farias Neto et al. (2007)
Oil content	2	C1	1	Fasoula et al. (2004)
	2	H	1	Fasoula et al. (2004)
	7	I	1	Nichols et al. (2006)
	2	L	1	Fasoula et al. (2004)
Protein content	2	E	1	Fasoula et al. (2004)
	10	I	1	Sebolt et al. (2000); Nichols et al. (2006); Bolon et al. (2010)
Seed weight	2	G	2	Fasoula et al. (2004)
	7	I	1	Nichols et al. (2006)
Seed yield	6	B2	1	Concibido et al. (2003)
	7	I	1	Nichols et al. (2006)
Pod maturity	7	I	1	Nichols et al. (2006)
Low cadmium accumulation	1	K	1	Jegadeesan et al. (2010)
Salt tolerance	3	N	1	Hamwieh et al. (2011)
Total			25	

backcross ($BC1F_2$) in the PI 567374 resistance source and in a population of F_8 plants derived from a heterozygous F_5 plant in the Ripley resistance source. The QTL was significant in confirmation populations in both resistant backgrounds.

Soybean seed is a major source of protein for animal feed and oil for human consumption. Since most experimental data show that protein and oil content are negatively correlated, therefore simultaneous increases in protein and oil content can proceed only to a limited extent. There are a number of QTL mapping studies for seed protein and oil content, however only two QTLs for seed protein content and four QTLs for oil content have been confirmed so far (Sebolt et al. 2000; Fasoula et al. 2004; Nichols et al. 2006; Bolon et al. 2010). Seed weight, is an important yield component of soybean and is positively correlated with seed yield. Three QTLs for seed weight and two QTL for seed yield have been confirmed (Fasoula et al. 2004; Concibido et al. 2003; Nichols et al. 2006). Although seed yield is considered the trait of highest priority but its low heritability, requirement for extensive data collection

across environment and expectant large number of QTL with majority conditioning small effects, has most likely limited the number of QTLs identified and validated. One QTL for pod maturity and one QTL for low cadmium accumulation have been also validated in soybean (Nichols et al. 2006; Jegadeesan et al. 2010). Recently, Hamwieh et al. (2011) validated a major salt tolerant QTL on linkage group N, using three sets of near isogenic lines (NILs). The evaluation of salt tolerance of the NILs revealed that all the lines with salt tolerant parent chromosome segment at the QTL region showed significantly higher salt tolerance than the lines without the tolerant parent chromosome segment. Although the number of confirmed QTLs in soybean is very less, the list is expected to increase in near future because of the development of advanced and high throughput technologies in genomics.

Marker Assisted QTL Introgression and Gene Pyramiding

The availability of molecular markers makes it possible to identify specific genomic regions and transfer them into commercial varieties with minimal linkage drag. The use of markers in backcrossing is especially applicable in transferring, into elite breeding material, novel genes that were obtained through genetic transformation or from exotic germplasm (Ribaut and Hoisington 1998). An example of marker assisted QTL introgression from exotic germplasm in soybean is reported by Sebolt et al. (2000). Two major QTL alleles from *Glycine soja* that increased seed protein concentration were identified previously by Diers et al. (1992a), used for marker assisted introgression. The *G. soja* genomic segments on LG I, which had the greatest effect, was backcrossed into the background of the experimental line with the help of linked markers. The BC_3F_4-derived lines that were homozygous for the *G. soja* QTL allele had a 20 g kg^{-1} greater seed protein concentration than lines homozygous for the soybean allele. They crossed the *G. soja* QTL into the backgrounds of two high yielding cultivars and one high protein experimental line. They found a significant ($P<0.0001$) increase in protein concentration associated with the *G. soja* QTL in the former two populations but not the population derived from the high protein experimental line. These results suggest that the high protein QTL allele from *G. soja* that was studied could already exist in the high protein experimental line. This experiment demonstrate the usefulness of markers in introgressing genes from exotic sources but also the fact that "mined" alleles may already be present in elite germplasm. Concibido et al. (2003) identified a yield-enhancing QTL from *G. soja* by evaluating a population of 265 BC_2 individuals from a cross between HS-1 and PI 407305. The yield QTL was located on linkage group B2. To assess the adaptability of the *G. soja* yield-QTL across genetic backgrounds, they developed sibling lines derived from BC_1 and BC_2 populations in the following Asgrow germplasm: AG4501, AG2401, QR4459, QP4459, QR4544 and QP4604. Interestingly, the efficacy of the yield QTL was limited to AG4501 and QP4459 lines, indicating considerable effect of genetic background on QTL expression. Lines that were homozygous for the PI 407305 haplotype at the QTL locus demonstrated a 9 % yield increase ($P=0.0006$) over lines that were homozygous for the AG4501 haplotype, whereas, in the QP4459

background, the presence of the PI 407305 haplotype showed a 5 % yield increase ($P=0.0119$) over lines that were homozygous for the QP4459 haplotype.

Gene pyramiding has been used as an effective approach to achieve multiple and durable resistance in crop plants. To evaluate a multiple resistance gene pyramiding strategy, Walker et al. (2004) developed eight soybean lines possessing factorial combinations of two quantitative trait loci (QTLs) for insect resistance from plant introduction (PI) 229358 and a synthetic Bt *cry1Ac* gene using marker assisted selection with simple sequence repeat (SSR) markers. Field studies were conducted to evaluate resistance to corn earworm (*Helicoverpa zea*) and soybean looper (*Pseudoplusia includens*), and detached leaf bioassays were used to test antibiosis resistance to Bt-resistant and Bt-susceptible strains of tobacco budworm (TBW; *Heliothis virescens*). Based on defoliation in the field and larval weight gain on detached leaves, lines carrying a combination of *cry1Ac* and the PI 229358 allele at a QTL on linkage group M were significantly more resistant to the lepidopteran pests, including the Bt-resistant TBW strain, than were the other lines.

Gene pyramiding has been used to achieve multiple and durable resistance to various strains of Soybean Mosaic Virus (SMV) in soybean. Shi et al. (2009) have successfully pyramided three genes *Rsv1*, *Rsv3*, and *Rsv4* for SMV resistance with the help of SSR markers in order to develop new soybean lines harboring multiple resistance genes. Genotype J05 carrying *Rsv1* and *Rsv3* and V94-5152 carrying *Rsv4* were used as the donor parents for gene pyramiding. A series of $F_{2:3}$, $F_{3:4}$, and $F_{4:5}$ lines derived from J05 × V94-5152 were developed for selecting individuals carrying all three genes. Eight PCR-based markers linked to the three SMV resistance genes were used for marker-assisted selection. Two SSR markers (Sat_154 and Satt510) and one gene-specific marker (*Rsv1-f/r*) were used for selecting plants containing *Rsv1*; two SSR markers (Satt560 and Satt063) for *Rsv3*; and three (Satt266, AI856415, and AI856415-g) for *Rsv4*. Five $F_{4:5}$ lines were homozygous for all eight marker alleles and presumably carry all three SMV resistance genes that would potentially provide multiple and durable resistance to SMV. To evaluate the effects of pyramided QTLs on the level of tolerance to *Phytophthora* root rot (PRR) of soybean, Li et al. (2010) stacked seven QTLs underlying tolerance to PRR, and found that the accumulation of tolerance loci was positively correlated with decreases in disease loss percentage. The pyramid of loci underlying tolerance to PRR provided germplasm useful for crop improvement by marker-assisted selection and may provide durable cultivar tolerance against the PRR disease. In conclusion, molecular markers are valuable tool to accelerate and to improve the efficiency of breeding programmes for cultivar development in soybean.

Comparative and Functional Genomics

New short-read sequencing technologies, capable of producing vast quantities of sequence, have the potential to rapidly change the comparative sequence research landscape in the legumes. At the time of writing, some ongoing projects include genome resequencing, digital gene expression, and whole transcript profiling.

Projects are being carried out both in legumes with near-complete genome sequence, and in less well-developed research systems. Innes et al. (2008) sequenced an approximately 1 Mbp region in soybean centered on the *Rpg1b* disease resistance gene and compared this region with a region duplicated 10–14 Mya. These two regions were also compared with homologous regions in several related legume species (a second soybean genotype *G. tomentella* D3, *Phaseolus vulgaris*, and *M. truncatula*). In this study one hundred and ten BACs were sequenced from different legumes (http://sites.bio.indiana.edu/~nsflegume/progress.php). Comparison of ~1 Mb region of soybean with the orthologous regions of *Medicago*, *Phaseolus* and *G. tomentella* D3 addressed several fundamental questions relating to Nucleotide binding-Leucine rich repeat (NB-LRRs) genes, polyploidy, and genome evolution. Analysis revealed a high level of conservation of low-copy genes but major differences in the NB-LRR content and retroelement content as well as differences in copy number of a family of protein kinases. Ashfield et al. (2012) used a comparative genomics approach to investigate the evolution of a complex nucleotide-binding (NB)-leucine-rich repeat (LRR) gene cluster found in soybean that is associated with several disease resistance (R) genes of known function, including *Rpg1b* (for Resistance to *Pseudomonas glycinea1b*), an R gene effective against specific races of bacterial blight. Analysis of domains revealed that the amino-terminal coiled-coil (CC) domain, central nucleotide-binding domain (NB-ARC), and carboxyl-terminal LRR domain have undergone distinct evolutionary paths. Sequence exchanges within the NB-ARC domain were rare. In contrast, interparalogue exchanges involving the CC and LRR domains were common, consistent with both of these regions coevolving with pathogens.

Retrotransposons and their remnants often constitute more than 50 % of higher plant genomes. Although extensively studied in monocot crops such as maize and rice, the impact of retrotransposons on dicot crop genomes is not well documented. Wawrzynski et al. (2008) identified several retrotransposon families in the soybean and grouped the 23 intact elements into 16 families. Nine of these 16 families contain elements that had inserted within the last million years, and two elements contained identical long terminal repeats (LTRs). In addition to this, several apparently replicating non-autonomous retrotransposon families were identified. The results indicated that autonomous and non-autonomous retro-transposons appear to be both abundant and active in the soybean. Gill et al. (2009) characterized and analyzed two subfamilies of high-copy centromeric satellite repeats, CentGm-1 and CentGm-2, using a combination of computational and molecular cytogenetic approaches. These two subfamilies of satellite repeats mark distinct subsets of soybean centromeres and, in at least one case, a pair of homoeologs, suggesting their origin from an allopolyploid event. These satellite repeats are also present in *G. soja*, the wild progenitor of soybean, but could not be detected in any other relatives of soybean examined in this study, suggesting the rapid divergence and species-specific concerted evolution of the centromeric satellite DNA within the *Glycine* genus. Recently Tian et al. (2012) investigated TE insertions in 31 resequenced wild and cultivated soybean genomes and detected 34,154 unique nonreference TE insertions mappable to the reference genome. Data revealed consistent distribution patterns of the nonreference LTR-RT insertions and those present in the reference genome, whereas the distribution patterns of the nonreference DNA TE insertions and the accumulated ones were significantly different.

Gene expression studies are imperative constituent of any crop improvement programme and soybean is no different. Expression studies on single genes employing RNA blotting or quantitative RT PCR etc were in vogue however global gene expression pattern analysis forms an integral part of soybean functional genomics. The gene expression patterns are investigated using the global expression analysis techniques like high-density expression arrays with clones imprinted on conventional nylon filters detected using radioactive probes, micro array systems with cDNA clones imprinted on glass slides with fluorescently labeled probe detection and finally Serial Analysis of Gene Expression (SAGE) meant for both quantitative and qualitative gene expression analysis. Microarray based expression investigation on soybean was initiated with a mere 18,000 cDNAs arrayed on a filter in high density expression arrays format (Vodkin et al. 2004). Later on the usage of microarray on soybean gene expression studies were very sparse like the instances of comparison of gene expression between root and shoot (Maguire et al. 2002), comparing transcript expression pattern during somatic embryogenesis (Thibaud-Nissen et al. 2003). Similarly to unravel the genes responsible for SCN susceptibility, soybean cv Kent was diagnosed using microarrays containing over 1,300 cDNA inserts specifically isolated from soybean libraries infected with SCN population (Khan et al. 2004). It was followed by the array of 27,513 cDNA based, low redundancy unigene sets representing wide range of source tissue and organ systems, developmental stages, and stress or pathogen-challenged plants (Vodkin et al. 2004). Further the quality of the cDNA microarray was scrutinized and demonstrated for its competence to distinguish the isogenic, mutant lines which are differentiated by expression or otherwise of small list of candidate genes. Currently soybean genome arrays have been designed and developed in consultation with Soybean Research Community in a consortia mode by Affymetrix. GeneChip® Soybean Genome Array by Affymetrix is characterized by the utility on expression studies of over 37,500 transcripts. The array also includes probe sets to detect over 15,000 and 7,500 transcripts of *Phytophthora sojae* and *Heterodera glycines* transcripts, respectively.

MicroRNAs (miRNAs) are key regulators of gene expression and play important roles in many aspects of plant biology. Turner et al. (2012), identified number of novel miRNAs and previously unknown family members for conserved miRNAs in the recently released soybean genome sequence. They classified all known soybean miRNAs based on their phylogenetic conservation (conserved, legume- and soybean-specific miRNAs) and examined their genome organization, family characteristics and target diversity. Comparative and functional genomics of soybean has covered in great detail by Ma et al. (2010), Livingstone et al. (2010).

Conclusions

In just the past few years we have witnessed tremendous progress in soybean genomics and an explosive expansion of new resources. We have seen the development of high-density genetic maps, construction of physical and transcript maps, EST

sequencing and analysis, development of high-density cDNA and oligo arrays, and sequencing and comparison of homologous segments. These resources and the resultant studies have shed much light on the structure, organization and evolution of the soybean genome. With the availability of the whole-genome sequence of the soybean genome, large-scale genomic sequences from two other reference legume species, *M. truncatula* and *L. japonicus,* plus the emerging genomic data from other legumes (Pigeonpea and Common bean), multi-species genome-wide comparisons can be achieved. These approaches will allow researchers to decipher the evolutionary history and genomic complexity of legumes. We will be able to further explore genomic approaches to the elucidation of key genes or functional components that control complex agronomical and physiological traits.

References

Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131–1139

Akkaya MS, Shoemaker RC, Specht JE, Bhagwat AA, Cregan PB (1995) Integration of simple sequence repeat (SSR) DNA markers into a soybean linkage map. Crop Sci 35:1439–1445

Alkharouf N, Matthews BF (2004) SGMD: the soybean genomics and microarray database. Nucleic Acids Res 32:1e3

Apuya NR, Frazier BL, Keim P, Roth EJ, Lark KG (1988) Restriction fragment length polymorphisms as genetic markers in soybean, Glycine max (L.) Merrill. Theor Appl Genet 75: 889–901

Arumuganathan K, Earle ED (1991) Estimation of nuclear DNA content of plants by flow cytometry. Plant Mol Biol Rep 9:229–241

Ashfield T, Egan AN, Pfeil BE, Chen NW, Podicheti R, Ratnaparkhe MB, Ameline-Torregrosa C, Denny R, Cannon S, Doyle JJ et al (2012) Evolution of a complex disease resistance gene cluster in diploid *Phaseolus* and tetraploid *Glycine*. Plant Physiol 159:336–354

Bolon YT, Joseph B, Cannon SB, Graham MA, Diers BW, Farmer AD, May GD, Muehlbauer GJ, Specht JE, Tu ZJ, Weeks N, Xu WW, Shoemaker RC, Vance CP (2010) Complementary genetic and genomic approaches help characterize the linkage group I seed protein QTL in soybean. BMC Plant Biol 10:41

Brown-Guedira GL, Thompson JA, Nelson RL, Warburton ML (2000) Evaluation of genetic diversity of soybean introductions and north American ancestors using RAPD and SSR markers. Crop Sci 40:815–823

Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16:1667–1678

Chaisan T, Van K, Kim MY, Kim KD, Choi BS, Lee SH (2010) *In silico* single nucleotide polymorphism discovery and application to marker-assisted selection in soybean. Mol Breed. doi:10.1007/s11032-010-9541-y

Choi IY, Hyten DL, Lakshmi KM, Qijian S, Julian MC, Charles VQ, Kevin C, Lark KG, Robert SR, Mun SY, Eun YH, Seung IY, Nevin DY, Randy CS, Curtis PVT, James ES, Cregan PB (2007) A soybean transcript map: gene distribution, haplotype and SNP analysis. Genetics 176:685–696

Concibido VC, La Vallee B, Mclaird P, Pineda N, Meyer J, Hummel L, Yang J, Wu K, Delannay X (2003) Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. Theor Appl Genet 106:575–582

Cregan PB (2008) The soybean molecular genetic linkage map. In: Stacy G (ed) Genetics and genomics of soybean. Springer Science+Business Media, LLC, New York, NY, USA, pp 79–89

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, VanToai TT, Lohnes DG, Chung J, Specht JE (1999) An integrated genetic linkage map of the soybean. Crop Sci 39:1464–1490

Danesh D, Penula S, Mudge J, Denny RL, Nordstrom H, Martinez JP, Young ND (1998) A bacterial artificial chromosome library for soybean and identification of clones near a major cyst nematode resistance gene. Theor Appl Genet 96:196–202

Darlington CD, Wylie AP (1955) Chromosome atlas of flowering plants. Allen and Unwin, London, UK

De Vries B (2011) Impact of the MON89788 event for glyphosate tolerance on agronomic and seed traits of soybean and molecular characterization of the mutant *fap3*(A22) allele for reduced palmitate concentration in soybean. Graduate Theses and Dissertations. Paper 11910. http://lib.dr.iastate.edu/etd/11910. Last Accessed on 26 Feb 2013

Diers BW, Keim P, Fehr WR, Shoemaker RC (1992a) RFLP analysis of soybean seed protein and oil content. Theor Appl Genet 83:608–612

Diers BW, Mansur L, Imsande J, Shoemaker RC (1992b) Mapping of *Phytophthora* resistance loci in soybean with restriction fragment length polymorphism markers. Crop Sci 32:377–383

Dietrich WF, Miller JC, Steen RG, Merchant M, Damron D, Nahf R, Gross A, Joyce DC, Wessel M, Dredge RD, Marquis A, Stein LD, Goodman N, Page DC, Lander ES (1994) A genetic map of the mouse with 4006 simple sequence length polymorphism. Nature Genet 7:220–225

Du JC, Grant D, Tian ZX, Nelson RT, Zhu LC, Shoemaker RC, Ma JX (2010) SoyTEdb: a comprehensive database of transposable elements in the soybean genome. BMC Genomics 11:113

Farias Neto ALD, Hashmi R, Schmidt M, Carlson SR, Hartman GL, Li S, Nelson RL, Diers BW (2007) Mapping and confirmation of a new sudden death syndrome resistance QTL on linkage group D2 from the soybean genotypes PI 567374 and "Ripley". Mol Breed 20:53–62

Fasoula VA, Harris DK, Boerma HR (2004) Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. Crop Sci 44: 1218–1225

Ferreira AR, Foutz KR, Keim P (2000) Soybean genetic map of RAPD markers assigned to an existing scaffold RFLP map. J Hered 91:392–396

Gill N, Findley S, Walling JG, Ma J, Stacey G, Doyle J, Jackson SA (2009) Molecular and chromosomal evidence for allo-polyploidy in soybean *Glycine max* (L) Merr. Plant Physiol 151:1167–1174

Gillman JD, Pantalone VR, Bilyeu K (2009) The low phytic acid phenotype in soybean line CX1834 is due to mutations in two homologs of the maize low phytic acid gene. J Plant Genom 2:179–190

Glover KD, Wang D, Arelli PR, Carlson SR, Cianzio SR, Diers BW (2004) Near isogenic lines confirm a soybean cyst nematode resistance gene from PI 88788 on linkage group. J Crop Sci 44:936–941

Goldblatt P (1981) Cytology and the phylogeny of Legumminosae. In: Polhill RM, Raven PH (eds) Advances in legume systematics, part II. Royal Botanic Garden, Kew, UK, pp 427–463

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178–D1186

Grant D, Cregan P, Shoemaker RC (2000) Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. Proc Natl Acad Sci U S A 97:4168–4173

Grant D, Nelson RT, Cannon SB, Shoemaker RC (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Res 38:D843–D846

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, et al (2003) Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res 31: 5654–5666

Hamilton C, Frarey A, Lewis C, Tanksley SD (1996) Stable transfer of intact high molecular weight DNA into plant chromosomes. Proc Natl Acad Sci U S A 93:9975–9979

Hamwieh A, Tuyen DD, Cong H, Benitez ER, Takahashi R, Xu DH (2011) Identification and validation of a major QTL for salt tolerance in soybean. Euphytica 179(3):451–459

Hisano H, Sato S, Isobe S, Sasamoto S, Wada T, Matsuno A, Fujishiro T, Yamada M, Nakayama S, Nakamura Y, Watanabe S, Harada K, Tabata S (2007) Characterization of the soybean genome using EST-derived microsatellite markers. DNA Res 14:271–281

Hwang TY, Sayama T, Takahashi M, Takada Y et al (2009) High-density integrated linkage map based on SSR markers in soybean. DNA Res 16:213–225

Hyten DL, Cannon SB, Song Q, Weeks N, Fickus EW, Shoemaker RC, Specht JE, Farmer AD, May GD, Cregan PB (2010a) High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. BMC Genomics 11:38

Hyten DL, Choi I-Y, Song Q, Specht JE, Carter TE, Shoemaker RC, Hwang EY, Matukumalli LK, Cregan PB (2010b) A high density integrated genetic linkage map of soybean and the development of a 1,536 Universal Soy Linkage Panel for QTL mapping. Crop Sci 50:960–968

Hyten DL, Song Q, Choi IY, Yoon MS, Specht JE, Matukumalli LK, Nelson RL, Shoemaker RC, Young ND, Cregan PB (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor Appl Genet 116:945–952

Innes RW, Ameline-Torregrosa C, Ashfield T, Cannon E, Cannon SB, Chacko B, Chen NW, Couloux A, Dalwani A, Denny R, Deshpande S, Egan AN, Glover N, Hans CS, Howell S, Ilut D, Jackson S, Lai H, Mammadov J, Del Campo SM, Metcalf M, Nguyen A, O'Bleness M, Pfeil BE, Podicheti R, Ratnaparkhe MB, Samain S, Sanders I, Segurens B, Sevignac M, Sherman-Broyles S, Thareau V, Tucker DM, Walling J, Wawrzynski A, Yi J, Doyle JJ, Geffroy V, Roe BA, Maroof MA, Young ND (2008) Differential accumulation of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean. Plant Physiol 148:1740–1759

Jegadeesan S, Yu K, Poysa V, Gawalko E, Morrison MJ, Shi C, Cober E (2010) Mapping and validation of simple sequence repeat markers linked to a major gene controlling seed cadmium accumulation in soybean [*Glycine max* (L.) Merr]. Theor Appl Genet 121:283–294

Joshi T, Patil K, Fitzpatrick MR, Franklin LD, Yao Q, Cook JR, Wang Z, Libault M, Brechenmacher L, Valliyodan B, Wu X, Cheng J, Stacey G, Nguyen HT, Xu D (2012) Soybean Knowledge Base (SoyKB): a web resource for soybean translational genomics. BMC Genomics 13(Suppl 1):S15

Juwattanasomran R, Somta P, Chankaew S, Shimizu T, Wongpornchai S, Kaga A, Srinives P (2011) A SNP in *GmBADH2* gene associates with fragrance in vegetable soybean variety "Kaori" and SNAP marker development for the fragrance. Theor Appl Genet 122:533–541

Juwattanasomran R, Somta P, Kaga A, Chankaew S, Shimizu T, Sorajjapinun W, Srinives P (2012) Identification of a new fragrance allele in soybean and development of its functional marker. Mol Breed 29:13–21

Kabelka EA, Carlson SR, Diers BW (2005) Localization of two loci that confer resistance to soybean cyst nematode from *Glycine soja* PI 468916. Crop Sci 45:2473–2481

Keim P, Beavis W, Schupp J, Freestone R (1992) Evaluation of soybean RFLP marker diversity in adapted germplasm. Theor Appl Genet 85:205–212

Keim P, Diers BW, Olson TC, Shoemaker RC (1990) RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics 126:735–742

Keim P, Schupp JM, Travis SE, Clayton K, Zhu T, Shi L, Ferreira A, Webb DM (1997) A high-density soybean genetic map based on AFLP markers. Crop Sci 37:537–543

Keim P, Shoemaker RC, Palmer RG (1989) Restriction fragment length polymorphism diversity in soybean. Theor Appl Genet 77:786–792

Khan R, Alkharouf N, Beard H, MacDonald M, Chouikha I, Meyer S, Grefenstette J, Knap H, Matthews B (2004) Microarray analysis of gene expression in soybean roots susceptible to the soybean cyst nematode two days post invasion. J Nematol 36(3):241–248

Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 39(Database issue):D152–D157

Lam HM, Xu X, Liu X, Chen WB, Yang GH, Wong FL, Li MW, He WM, Qin N, Wang B, Li J, Jian M, Wang JA, Shao G, Wang J, Sun SSM, Zhang GY (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat Genet 42:1053–1059

Lark KG, Weisemann JM, Matthews BF, Palmer R, Chase K, Macalma T (1993) A genetic map of soybean (Glycine max L.) using an intraspecific cross of two cultivars: "Minsoy" and "Noir 1". Theor Appl Genet 86:901–906

Lenis JM, Gillman JD, Lee JD, Shannon JG, Bilyeu KD (2010) Soybean seed lipoxygenase genes: molecular characterization and development of molecular marker assays. Theor Appl Genet 120:1139–1149

Li X, Han Y, Teng W, Zhang S, Yu K, Poysa V, Anderson T, Ding J, Li W (2010) Pyramided QTL underlying tolerance to Phytophthora root rot in mega-environments from soybean cultivars "Conrad" and "Hefeng 25". Theor Appl Genet 121:651–658

Li Z, Nelson RL (2002) RAPD marker diversity among cultivated and wild soybean accessions from four Chinese provinces. Crop Sci 42:1737–1744

Liu F, Dong FY, Zou JJ, Chen SY, Zhuang BC (2000) Soybean germplasm diversity and genetic variation detected by microsatellite markers. Chinese Acta Genet Sin 27:628–633

Lightfoot DA, Njiti VN, Gibson PT, Kassem MA, Iqbal JM, Meksem K (2005) Registration of the Essex by Forrest recombinant inbred line mapping population. Crop Sci 45:1678–1681

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397–401

Livingstone JM, Cheng KC, Strömvik MV (2010) Bioinformatics as a tool. In: Bilyeu K, Ratnaparkhe MB, Kole C (eds) Genetics, genomics, and breeding of soybean. CRC, New Hampshire, UK

Lorenzen LL, Boutin S, Young N, Specht JE, Shoemaker RC (1995) Soybean pedigree analysis using map-based molecular markers: I. Tracking RFLP markers in cultivars. Crop Sci 35: 1326–1336

Luo MC, Thomas C, You FM, Hsiao J, Ouyang S, Buell CR et al (2003) High-throughput fingerprinting of bacterial artificial chromosomes using the SNaPshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82:378–389

Ma J, Shoemaker R, Jackson S, Cannon S (2010) Comparative genomics. In: Bilyeu K, Ratnaparkhe MB, Kole C (eds) Genetics, genomics, and breeding of soybean. CRC, New Hampshire, UK

Maguire TL, Grimmond S, Forrest A, Iturbe-Ormaetze I, Meksem K, Gresshoff P (2002) Tissue-specific gene expression in soybean (*Glycine max*) detected by cDNA microarray analysis. Plant Physiol 159:1361–1374

Mansur LM, Orf JH, Chase K, Jarvik T, Cregan PB, Lark KG (1996) Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Sci 36:1327–1336

Marek LF, Shoemaker RC (1997) BAC contig development by fingerprint analysis in soybean. Genome 40:420–427

Matthews BF, Devine TE, Weisemann JM, Beard HS, Lewers KS, MacDonald MH, Park YB, Maiti R, Lin JJ, Kuo J (2001) Incorporation of sequenced cDNA and genomic markers into the soybean genetic map. Crop Sci 41:516–521

Matthews BF, MacDonald MH, Gebhardt JST, Devine TE (1998) Molecular markers residing close to the Rhg4 locus conferring resistance to soybean cyst nematode race 3 on linkage group A of soybean. Theor Appl Genet 97:1047–1052

Maughan PJ, Saghai Maroof MA, Buss GR (1995) Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. Genome 38:715–723

Meksem K, Pantazopoulos P, Njiti VN, Hyten LD, Arelli PR, Lightfoot DA (2001) "Forrest" resistance to the soybean cyst nematode is bigenic: saturation mapping of the Rhg1 and Rhg4 loci. Theor Appl Genet 103:710–717

Meksem K, Zobrist K, Ruben E, Hyten D, Quanzhou T, Zhang HB, Lightfoot DA (2000a) Two large-insert soybean genomic libraries constructed in a binary vector: applications in chromosome walking and genome wide physical mapping. Theor Appl Genet 101:747–755

Meksem K, Ruben E, Zobrist K, Zhang HB, Lightfoot D (2000b) Two large insert libraries for soybean: application in cyst nematode resistance and genome wide physical mapping. Theor Appl Genet 101:747–755

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3:175–182

Morgante M, Rafalski A, Biddle P, Tngey S, Olivery AM (1994) Genetic mapping and variability of seven soybean simple sequence repeat loci. Genome 37:763–769

Nichols DM, Glover KD, Carlson SR, Specht JE, Diers BW (2006) Fine mapping of a seed protein qtl on soybean linkage group I and its correlated effects on agronomic traits. Crop Sci 46:834–839

Ratnaparkhe MB, Singh RJ, Doyle JJ (2011) *Glycine*. In: Kole C (ed) Wild crop relatives: genomic and breeding resources. Springer, Heidelberg, Germany, pp 83–116

Prabhu RR, Njiti VN, Bell-Johnson B, Johnson JE, Schmidt ME, Klein JH, Lightfoot DA (1999) Selecting soybean cultivars for dual resistance to soybean cyst nematode and sudden death syndrome using two DNA markers. Crop Sci 39:982–987

Ribaut J, Hoisington D (1998) Marker-assisted selection: new tools and strategies. Trends Plant Sci 3:236–239

Rongwen J, Akkaya MS, Bhagwat AA, Lavi U, Cregan PB (1995) The use of microsatellite DNA markers for soybean genotype identification. Theor Appl Genet 90:43–48

Salimath SS, Bhattacharyya MK (1999) Generation of a soybean BAC library, and identification of DNA sequences tightly linked to the Rps1-k disease resistance gene. Theor Appl Genet 98:712–720

Schmutz J, Cannon SB, Schlueter J, Ma JX, Mitros T, Nelson W, Hyten DL, Song QJ, Thelen JJ, Cheng JL, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu SQ, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du JC, Tian ZX, Zhu LC, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Schuler GD (1997) Pieces of the puzzle: expressed sequence tags and the catalog of human genes. J Mol Med 75:694–698

Sebolt AM, Shoemaker RC, Diers BW (2000) Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. Crop Sci 40:1438–1444

Sha AH, Li C, Yan XH, Shan ZH, Zhou XA, Jiang ML, Mao H, Chen B, Wan X, Wei WH (2012) Large-scale sequencing of normalized full-length cDNA library of soybean seed at different developmental stages and analysis of the gene expression profiles based on ESTs. Mol Biol Rep 39:2867–2874

Shi A, Chen P, Li D, Zheng C, Zhang B, Hou A (2009) Pyramiding multiple genes for resistance to soybean mosaic virus in soybean using molecular markers. Mol Breed 23:113–124

Shoemaker RC, Grant D, Olson T, Warren WC, Wing R, Yu Y, Kim H, Cregan P, Joseph B, Futrell-Griggs M, Nelson W, Davito J, Walker J, Wallis J, Kremitski C, Scheer D, Clifton SW, Graves T, Nguyen H, Wu X, Luo M, Dvorak J, Nelson R, Cannon S, Tomkins J, Schmutz J, Stacey G, Jackson S (2008) Microsatellite discovery from BAC end sequences and genetic mapping to anchor the soybean physical and genetic maps. Genome 51:294–302

Shoemaker RC, Specht JE (1995) Integration of the soybean molecular and classical genetic linkage groups. Crop Sci 35:436–446

Shoemaker RC, Vodkin L, Schlueter JA, Cregan P (2003) The status of soybean genomics and its role in the development of soybean biotechnologies. AgBioForum 6:4–7

Shoemaker R, Keim P, Vodkin L, Retzel E, Clifton SW, Waterston R, Smoller D, Coryell V, Khanna A, Erpelding J, Gai X, Brendel V, Raph-Schmidt C, Shoop EG, Vielweber CJ, Schmatz M, Pape D, Bowers Y, Theising B, Martin J, Dante M, Wylie T, Granger C (2002) A compilation of soybean ESTs: generation and analysis. Genome 45:329–338

Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP, Kochert G, Boerma HR (1996) Genome duplication in soybean (Glycine subgenus soja). Genetics 144:329–338

Shoemaker RC, Schlueter J, Doyle JJ (2006) Paleopolyploidy and gene duplication in soybean and other legumes. Curr Opin Plant Biol 9:104–109

Shultz JL, Kazi S, Bashir R, Afzal JA, Lightfoot DA (2007) The development of BAC-end sequence-based microsatellite markers and placement in the physical and genetic maps of soybean. Theor Appl Genet 114:1081–1090

Singh RJ, Hymowitz T (1988) The genomic relationship between *Glycine max* (L.) Merr. and *G. soja* Sieb. and Zucc. as revealed by pachytene chromosome analysis. Theor Appl Genet 76:705–711

Skorupska HT, Shoemaker RC, Warner A, Shipe ER, Bridges WC (1993) Restriction fragment length polymorphism in soybean germplasm of the southern USA. Crop Sci 33:1169–1176

Song Q, Jia G, Zhu Y, Grant D, Nelson RT, Hwang E, Hyten DL, Cregan PB (2010) Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. Crop Sci 50:1950–1960

Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. Theor Appl Genet 109:122–128

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17:6463–6471

Thibaud-Nissen F, Shealy RT, Khanna A, Vodkin LO (2003) Clustering of microarray data reveals transcript patterns associated with somatic embryogenesis in soybean. Plant Physiol 132:118–136

Thompson JA, Nelson RL, Vodkin LO (1998) Identification of diverse soybean germplasm using RAPD markers. Crop Sci 38:1348–1355

Tian AG, Wang J, Cui P, Han YJ, Xu H, Cong LJ, Huang XG, Wang XL, Jiao YZ, Wang BJ et al (2004) Characterization of soybean genomic features by analysis of its expressed sequence tags. Theor Appl Genet 108:903–913

Tian Z, Zhao M, She M, Du J, Cannon SB, Liu X, Xu X, Qi X, Li MW, Lam HMJ Ma (2012) Genome-wide characterization of nonreference transposons reveals evolutionary propensities of transposons in soybean. Plant Cell 24(11):4422–4436, doi: http://dx.doi.org/10.1105/tpc.112.103630

Tomkins JP, Frish DA, Byrum JR, Wing RA (2000) Construction and characterization of a BAC library for the soybean cultivar A3244. Soybean Genet Newsletter 27. http://www.soygenetics.org. Accessed 26 Feb 2013

Tomkins JP, Mahalingham R, Miller-Smith H, Goicoechea JL, Knapp HT, Wing RA (1999) A soybean bacterial artificial chromosome library for PI 437654 and the identification of clones associated with cyst nematode resistance. Plant Mol Biol 41:25–32

Turner M, Oliver Y, Subramanian S (2012) Genome organization and characteristics of soybean microRNAs. BMC Genomics 13:169

Umezawa T, Sakurai T, Totoki Y, Toyoda A, Seki M, Ishiwata A, Akiyama K, Kurotani A, Yoshida T, Mochida K, Kasuga M, Todaka D, Maruyama K, Nakashima K, Enju A, Mizukado S, Ahmed S, Yoshiwara K, Harada K, Tsubokura Y, Hayashi M, Sato S, Anai T, Ishimoto M, Funatsuki H, Teraishi M, Osaki M, Shinano T, Akashi R, Sakaki Y, Yamaguchi-Shinozaki K, Shinozaki K (2008) Sequencing and analysis of approximately 40,000 soybean cDNA clones from a full-length-enriched cDNA library. DNA Res 15:333–346

Varshney RK, Hoisington DA, Tyagi AK (2006) Advences in cereal genomics and applications in crop breeding. Trends Biotechnol 24:490–499

Vodkin LO, Khanna A, Shealy R, Clough S, Gonzalez O, Philip R, Zabala G, Thibaud-Nissen F, Sidarous M, Strömvik M, Shoop E, Schmidt C, Retzel E, Erpelding J, Shoemaker R, Rodriguez-Huete A, Polacco J, Coryell V, Keim P, Gong G, Liu L, Pardinas J, Schweitzer P (2004) Microarrays for global expression constructed with a low redundancy set of 27,500 sequenced cDNAs representing an array of developmental stages and physiological conditions of the soybean plant. BMC Genomics 5:73

Vuong TD, Sleper DA, Shannon JG, Nguyen HT (2010) Novel quantitative trait loci for broad-based resistance to soybean cyst nematode (*Heterodera glycines* Ichinohe) in soybean PI 567516C. Theor Appl Genet 121:1253–1266

Walker DR, Narvel JM, Boerma HR, All JN, Parrott WA (2004) A QTL that enhances and broadens Bt insect resistance in soybean. Theor Appl Genet 109:1051–1057

Wang Z, Libault M, Joshi T, Valliyodan B, Nguyen HT, Xu D, Stacey G, Cheng JL (2010) SoyDB: a knowledge database of soybean transcription factors. BMC Plant Biol 10:14

Wawrzynski A, Ashfield T, Chen NW, Mammadov J, Nguyen A, Podicheti R, Cannon SB, Thareau V, Ameline-Torregrosa C, Cannon E, Chacko B, Couloux A, Dalwani A, Denny R, Deshpande S, Egan AN, Glover N, Howell S, Ilut D, Lai H, Del Campo SM, Metcalf M, O'Bleness M, Pfeil BE, Ratnaparkhe MB, Samain S, Sanders I, Segurens B, Sherman-Broyles S, Sevignac M, Tucker DM, Yi J, Doyle JJ, Geffroy V, Roe BA, Maroof MA, Young ND, Innes RW (2008) Replication of nonautonomous retroelements in soybean appears to be both recent and common. Plant Physiol 148:1760–1771

Warren WC and The Soybean Mapping Consortium (2006) A physical map of the "Williams 82" soybean (*Glycine max*) genome. Abstract W151. In: Plant and animal genomes XIV conference, San Diego, CA, USA, 14–18 Jan 2006

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396

Wu CC, Sun SK, Nimmakayala P, Santos FA, Meksem K, Springman R, Ding K, Lightfoot DA, Zhang HB (2004a) A BAC and BIBAC-based physical map of the soybean genome. Genome Res 14:319–326

Wu XL, Zhong GH, Findley SD, Cregan P, Stacey G, Nguyen HT (2008) Genetic marker anchoring by six-dimensional pools for development of a soybean physical map. BMC Genomics 9:28

Wu C, Springman R, Santos FA, Nimmakayala P, Meksem K, Scheuring C, Lightfoot DA, Zhang HB (2004b) Construction and characterization of a soybean bacterial artificial chromosome library and use of multiple complementary libraries for genome physical mapping. Theor Appl Genet 109:1041–1050

Xia Z, Sato H, Watanabe S, Kawasaki S, Harada K (2005) Construction and characterization of a BAC library of soybean. Euphytica 141:129–137

Yamanaka N, Ninomiya S, Hoshi M, Tsubokura Y, Yano M, Nagamura Y, Sasaki T, Harada K (2001) An informative linkage map of soybean reveals QTLs for flowering time, leaflet morphology and regions of segregation distortion. DNA Res 8:61–72

Yan HH, Mudge J, Kim DJ, Larsen D, Shoemaker RC, Cook DR, Young ND (2003) Estimates of conserved microsynteny among the genomes of *Glycine max*, *Medicago truncatula* and *Arabidopsis thaliana*. Theor Appl Genet 106:1256–1265

Zhang G, Gu C, Wang D (2009) Molecular mapping of soybean aphid resistance in PI 567541B. Theor Appl Genet 118:473–482

Zhang G, Gu C, Wang D (2010) A novel locus for soybean aphid resistance. Theor Appl Genet 120:1183–1191

Zhu S, Saski CA, Boerma HR, Tomkins JP, All JN, Parrott WA (2009) Construction of a bac library for a defoliating insect-resistant soybean and identification of candidate clones using a novel approach. Plant Mol Biol Rep 27:229–235

Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003) Single-nucleotide polymorphisms in soybean. Genetics 163:1123–1134

Zhu T, Shi I, Gresshoff P, Keim P (1996) Characterization and application of soybean YACs to molecular cytogenetics. Mol Gen Genet 252:483–488

Chapter 4
Advances in Chickpea Genomics

Pooran M. Gaur, Mahendar Thudi, Srinivasan Samineni, and Rajeev K. Varshney

Abstract Chickpea (*Cicer arietinum* L.), the second largest consumed pulse crop of the world after common bean, is grown in over 50 countries and traded across 140 countries. After several decades of slow progress, the recent years have witnessed spectacular progress in development of genetic (mapping populations) and genomic resources (structural and functional molecular markers, integrated genetic map and mapping of genes/quantitative trait loci; QTLs) in this crop. QTLs associated with traits of interest including resistance against wilt, *Ascochyta* blight, *Botrytis* grey mould and rust; tolerance against salinity and drought and agronomic traits have been identified and validated. A more than 30 genetic linkage maps available in this crop are useful resources for genetic analysis and marker assisted breeding. Genomic tools like bacterial artificial chromosome (BAC) libraries, expressed sequence tags (ESTs) and targeting induced local lesions in genome (TILLING) mutants have been developed in chickpea to facilitate the genome sequencing efforts in this crop. A major landmark in chickpea genomics has been the publication of 738 Mb draft whole genome sequence assembly of a kabuli variety, CDC Frontier. Now, chickpea is one of the most advanced grain legumes in terms of availability of genomic resources. Efforts have already begun on application of these genomics resources in chickpea improvement. This book chapter provides an update on the development of genetic and genomic resources for chickpea and their current and potential uses in chickpea improvement.

Keywords *Cicer arietinum* • Molecular markers • Genome sequence • Quantitative trait loci • Marker assisted breeding

P.M. Gaur, Ph.D. (✉) • M. Thudi, Ph.D. • S. Samineni, Ph.D. • R.K. Varshney, Ph.D.
Grain Legumes Research Program, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Hyderabad, Andhra Pradesh 502324, India
e-mail: p.gaur@cgiar.org

S. Gupta et al. (eds.), *Legumes in the Omic Era*, DOI 10.1007/978-1-4614-8370-0_4,

Introduction

Chickpea (*Cicer arietinum* L.) is the second most important pulse crop of the world in terms of area and production. During 2010, chickpea was grown in more than 50 countries and had an area of about 12 million ha, production of 11 million tons and productivity of 910 kg ha^{-1} (FAOSTAT 2012). The major chickpea producing countries include India, Pakistan, Australia, Myanmar, Iran, Mexico, Canada and USA. The highest production and consumption of chickpea is in South Asia where India alone accounts for over two-third of the global area, production and consumption. The awareness of health benefits of chickpea has led to considerable increase in the international trade of chickpea.

Being a legume crop, chickpea is highly valued in the cropping systems, particularly in rotation with cereals, for its overall impacts on soil health. There has been a large shift in chickpea area (about 3 million ha) from cooler, long growing season environments to warmer, short growing season environments during the past four decades (Gaur et al. 2012a). This significant change in the chickpea growing environment and the expected impacts of climate change need to be accounted by chickpea breeding programs.

The major adaptation traits to be considered by chickpea breeding programs include phenology, plant type and resistance to key abiotic and biotic stresses prevalent in the target environment and growing conditions. Drought and heat stresses during the reproductive phase and with increasing severity towards the end of the crop season are the major abiotic stresses of chickpea as the crop is generally grown rainfed on residual soil moisture and experiences progressively receding soil moisture conditions and increasing atmospheric temperatures towards end of the crop season. Soil salinity and chilling atmospheric temperatures are also important stresses in some growing environments. Among diseases, fusarium wilt (caused by *Fusarium oxysporum* f. sp. *ciceri*), dry root rot (caused by *Rhizoctonia bataticola*), and collar rot (caused by *Sclerotium rolfsi*), are the important root diseases of chickpea in areas where the growing season is dry and warm, while ascochyta blight [caused by *Ascochyta rabiei* (Pass.) Labr.], and botrytis grey mold (caused by *Botrytis cineria* Pres.), are the important foliar diseases in the areas where the growing season is cool and humid. Pod borer (*Helicoverpa armigera* Hübner) is the most important pest of chickpea worldwide. The viral diseases, rust (caused by *Uromyces cicerisarietini*), root nematodes (*Meloidogyne* sp.), Phytophthora root rot (caused by *Phytophthora medicaginis*), cutworm (*Agrotis* sp.) and leaf miner (*Liriomyza cicerina*) are also important in some chickpea growing areas.

Recent advances in the development of genomic resources have made it possible to use genomics-assisted breeding for improvement of chickpea. The breeding programs will have higher precision and efficiency and thus better equipped to rapidly develop cultivars better adapted to existing and evolving growing environments and with improved nutrition quality and grain traits required by the industry and the consumers. This chapter provides an update on the progress made in development and use of genomic resources in chickpea.

Gene Pool

Chickpea is a self-pollinated, annual, diploid (2x = 2n = 16), cool season food legume. It is considered to have originated in south-eastern Turkey and the adjoining northern region of Syria (Van der Maesen and Pundir 1987), because the proposed wild progenitor (*C. reticulatum*) of the chickpea and its other closely related wild species (*C. echinospermum*, *C. bijugum*) are found there. The genus *Cicer* includes 43 species, 9 of which are annual, 33 are perennial and 1 with unspecified life cycle (Van der Maesen and Pundir 1987). The species *C. arietinum* is the only cultivated species of this genus. Based on successes in interspecific crosses, *C. arietinum* has been placed in primary gene pool, *C. echinospermum* in the secondary gene pool and all the remaining species in the tertiary gene pool (Ahmad et al. 2005). The phylogenetic relationships among nine annual species have also been studied based on allozyme polymorphism (Kazan and Muehlbauer 1991; Ahmad et al. 1992; Labdi et al. 1996; Tayyar and Waines 1996) protein banding patterns of seeds (Ahmad and Slinkard 1992) and randomly amplified polymorphic DNA (RAPD) markers (Ahmad 1999). These studies have categorised the annual *Cicer* species into four phylogenetic groups. *C. arietinum*, *C. reticulatum* and *C. echinospermum* formed one group while *C. pinnatifidum*, *C. bijugum* and *C. judaicum* formed another group. *C. chorassanicum* was grouped with *C. yamashitae* whereas *C. cuneatum* showed the largest distance from *C. arietinum* and formed an independent group. Further, cultivated chickpea was found to be more closely related to *C. reticulatum* than *C. echinospermum*. These results were further supported by studies using molecular markers such as RAPD (Iruela et al. 2002; Sudupak et al. 2002), amplified fragment length polymorphism (AFLP) (Nguyen et al. 2004; Sudupak 2004) and simple sequence repeats (SSR) (Croser et al. 2003; Rao et al. 2006; Choudhary et al. 2012a). In the process of evolution, chickpea has emerged into two distinct types; small seeded dark colored *Desi* and large seeded, cream colored *Kabuli*. About 80 % of the chickpea area is under the *Desi* type and the remaining area under the *Kabuli* type.

Molecular diversity studies indicated that wild relatives of chickpea have high genetic diversity compared to its cultivated species *C. arietinum* and supports the conclusion that chickpea has a narrow genetic base (Nguyen et al. 2004; Choudhary et al. 2012a). These results indicate that the varieties currently under cultivation are closely related among themselves. Efforts should be made to widen the genetic base of the cultigen by exploiting wild species. The wild species also offer opportunities of bringing novel alleles for important traits, particularly resistance to abiotic and biotic stresses (Gaur et al. 2010).

Genome and Genome Size

Almost all *Cicer* species have $2n = 2x = 16$ chromosomes. The genome of size of cultivated chickpea (*Cicer arietinum* L.) is 738 Mbp that is only 1.5 times higher than *Medicago truncatula*. The chromosomes have been numbered from 1 to 8 in

order of decreasing size of the chromosomes and the size difference between pair one and pair eight has been found to be in the ratio of 3:1 (Ahmad and Godward 1980). Ahmad and Hymowitz (1993) recorded the total chromosome length at pachytene stage as 353.53 μm and also found that the chromosome size ranged from 30.53 to 58.05 μm. The chickpea chromosomes are small which makes the karyotype analysis difficult. The chickpea karyotype revealed from various cytological investigations has the following features: a pair of very long chromosomes, distinctly satellited and sub-metacentric; six pairs of metacentric to sub-metacentric chromosomes; and a pair of very short metacentric chromosomes (reviewed by Gupta and Bahl 1983). Both, spontaneous (Sen and Jana 1956) and induced (Ramanujam and Joshi 1941; Akhtar 1954; Sharma and Gupta 1982; Pundir and Mengesha 1983), autotetraploids have been reported in chickpea. Seed treatment with 0.1–0.25 % colchicine for 4 h has been found effective in inducing autotetraploidy and these autotetraploids predominantly show bivalent pairing and normal disjunction at anaphase I (Sharma and Gupta 1983).

Genetic and Genomic Resources

Genetic resources, which include mapping populations, genetic stocks and breeding materials, have been developed in chickpea for use in genetic studies and breeding programs. Further, during recent years, large scale genomic resources in the form of molecular markers, genetic linkage maps and quantitative trait loci (QTL) maps have been developed and made available to breeders for implementing integrated breeding approaches and developing cultivars more efficiently.

Mapping Populations

Development of appropriate mapping population is necessary for constructing a genetic linkage map and dissecting complex traits. The first step in producing a mapping population is selecting two genetically diverse parents for one or more traits of interest. Further the parents should be genetically divergent enough to exhibit sufficient polymorphism, and on the other hand they should not be too distant that causes sterility of the progenies and expresses high level of segregation distortion during linkage analysis. A range of populations including progenies from F_2 generation, backcross (BC), recombinant inbred lines (RILs), double haploids (DH) and near isogenic lines (NILs) have been used for genetic mapping in chickpea. F_2 populations are developed by self-pollinating F_1 hybrids derived by crossing two parents, while BC population is produced by crossing F_1 to one of the parents. By repeated backcrossing for at least six generations (BC_6) with the recipient or recurrent parent, more than 99 % of the genome can be recovered from the recurrent parent. Further selfing of selected individuals at BC_6F_1 or BC_7F_1 will produce lines

Table 4.1 List of chickpea RIL mapping populations developed and available at ICRISAT

RIL population	Cross	Generation	No. of RILs	Segregating traits
ICCRIL01	ICCV 2 × JG 62	F_{10+}	573	Fusarium wilt (FW) resistance, botrytis gray mold (BGM) resistance, *Helicoverpa* resistance, salinity tolerance
ICCRIL02	Annigeri × ICC 4958	F_{10+}	257	Root traits
ICCRIL03	ICC 4958 × ICC 1882	F_{10+}	264	Root traits
ICCRIL04	ICC 283 × ICC 8261	F_{10+}	281	Root traits
ICCRIL05	ICC 506-EB × Vijay	F_{10+}	328	*Helicoverpa* resistance
ICCRIL06	ICC 3137 × IG 72953	F_6	241	*Helicoverpa* resistance
ICCRIL07	ICC 995 × ICC 5912	F_{10+}	240	Protein content
ICCRIL08	ICC 6263 × ICC 1431	F_8	266	Salinity tolerance
ICCRIL09	ICCV 2 × JG 11	F_8	280	Salinity tolerance
ICCRIL10	JG 62 × ICCV 05530	F_{10+}	315	Ascochyta blight (AB), BGM and FW resistance
ICCRIL11	Pb 7 × ICCV 04516	F_8	127	AB resistance
ICCRIL12	ICC 4567 × ICC 15614	F_8	296	Heat tolerance
ICCRIL13	ICC 4567 × ICC 1356	F_8	291	Heat tolerance

that are homozygous for the target gene, which are considered to be nearly isogenic with the recipient parent (NILs). NILs are mainly generated for fine mapping of a QTL/genomic region of interest. DH populations are generally developed by chromosome doubling of haploids developed though anther culture (pollen or microspore culture) of F_1 plants. RILs are developed following single seed descent (SSD) advancement of F_2 plants by six or more generations and then developing single plant progenies. This process leads to lines that contain a different combination of linkage blocks from the original parents. Seed from RILs is predominantly homogeneous and abundant, so the seed can be sent to any lab interested in adding markers to an existing linkage map previously constructed with the RILs. Moreover, RILs can be grown in replicated trials at several locations and/or over several years making them ideal for QTL mapping. Similar types of inbred populations, such as doubled haploids, can also be used for linkage mapping with many of the same advantages of RILs. The RIL mapping populations of chickpea developed and available at ICRISAT are listed in Table 4.1.

For creating novel genetic variation and identification of useful allelic variants, a TILLING (Targeting Induced Local Lesions IN Genomes) population from chickpea accession ICC 4958 was developed at ICRISAT through mutagenesis by ethyl methane sulphonate (EMS). This population comprises of >5,000 M2 lines which are currently being used for allele mining for various agronomically important genes. A multi-parent advanced generation inter-cross (MAGIC) population has been used to develop over 1,200 lines at ICRISAT. The MAGIC population was developed from eight parents and includes cultivars and elite breeding lines from India and Africa. Twenty-eight two-way, 14 four-way and 7 eight-way crosses were made to develop

this MAGIC population. The MAGIC lines constitute a valuable genetic resource for trait mapping and gene discovery. In addition, these can be directly used as source material for development of improved cultivars (Gaur et al. 2012b).

Molecular Markers

The genomic resources being made available for chickpea breeding community have been reviewed from time to time (Gaur et al. 2012b; Varshney et al. 2010, 2012a; Upadhyaya et al. 2011). However, this chapter provides the latest developments as well as discusses the pros and cons of these marker resources in various genetic analyses. Based on the method of detection of the sequence variation, the molecular markers can be classified as hybridization based (PCR-independent), PCR dependent and micro-array based markers. Restriction fragment length polymorphism (RFLP) markers were the first hybridization based highly reproducible, co-dominant, locus specific markers employed for plant genome analysis during 1990s. The first genetic map constructed in chickpea using molecular markers included RFLP and RAPD markers along with isozyme markers (Simon and Muehlbauer 1997). Genetic diversity studies were also carried out using RFLP markers (Udupa et al. 1993) and microsatellite-derived RFLP markers (Sharma et al. 1995; Serret et al. 1997). These studies showed narrow genetic variability for restriction sites in the genome of cultivated chickpea. The PCR-based marker systems are of two types—(1) non-sequence specific markers which include RAPD and AFLP markers, and (2) sequence tagged PCR-based markers which include cleaved amplified polymorphic sequence (CAPS), sequence tagged site (STS) and SSR markers. Although RAPD markers were also employed to characterize germplasm (Ahmad 1999; Sudupak et al. 2002), these markers are not currently being preferred for any genetic analysis in chickpea owing to the dominant nature of inheritance and non-reproducibility of these markers. However, utility of RAPD markers can be enhanced by converting these into more reproducible informative marker types such as sequence characterized amplified regions (SCAR). To overcome the limitations of reproducibility associated with RAPD, AFLP marker system was developed by selective amplification of DNA fragments obtained by restriction enzyme digestion. AFLP markers have been used for genetic diversity estimation in cultivated chickpea and its wild relatives in order to discover the origin and history of chickpea (Nguyen et al. 2004; Talebi et al. 2008, 2009). However, the requirement of significant technical skills, laboratory facilities, financial resources and high quality genomic DNA for complete restriction, digestion and dominant inheritance has limited the use of AFLP markers.

PCR based CAPS markers are characterized by their co-dominant inheritance and locus specific nature which are useful for genotyping applications (Parsons and Heflich 1997; Weiland and Yu 2003). In chickpea, CAPS and derived CAPS (dCAPS) markers have been developed from bacterial artificial chromosome (BAC)-end sequences (Rajesh et al. 2008) and EST sequences (Varshney et al.

2007), and these markers were further integrated into composite genetic map of chickpea to study their association with disease resistance (Palomino et al. 2009).

Microsatellite markers are also known as simple sequence repeats (SSRs) or sequence tagged microsatellite site (STMS), constitute tandem repeats of 1–6 bp in length (Gupta and Varshney 2000) are advantageous over many other markers types as they are highly polymorphic and abundant, analytically simple and readily transferable (Weber 1990), and show co-dominance. In chickpea genome SSRs were found to be abundant and showed moderately high level of intra-specific polymorphism when compared to other marker types (Sharma et al. 1995). About 500 SSR markers were available for chickpea earlier (Huttel et al. 1999; Winter et al. 1999; Lichtenzveig et al. 2005) and were used for development of genetic map (Winter et al. 1999; Millan et al. 2006). Later, several studies reported the development of SSR markers using hybridization based microsatellite enrichment and BAC and BIBAC libraries in chickpea (Lichtenzveig et al. 2005; Choudhary et al. 2006). At ICRISAT, currently >2,000 SSR markers are available for utilization in chickpea crop improvement (Varshney et al. 2009; Nayak et al. 2010; Thudi et al. 2011; Gujaria et al. 2011).

Sequence Information

Recent advances in next generation sequencing (NGS) technologies have greatly facilitated the ability to sequence the genome and transcriptomes of several plant species (Thudi et al. 2012). In case of chickpea, as on 13th November 2012, 97,836 nucleotide sequences were available in the public domain (http://www.ncbi.nlm.nih.gov/nuccore?term=chickpea%20cicer) against only a limited number of ESTs (Varshney et al. 2009).

Functional Markers, ESTs, BAC Libraries

Molecular markers developed from genes/ESTs are referred as genic molecular markers (GMMs; (Gujaria et al. 2011)) or functional markers (Anderson and Lübberstedt 2003). Based on origin, genic markers are of two kinds: (a) markers that are derived from polymorphisms within genes are gene targeted markers (GTMs), these markers however not necessarily involved in phenotypic trait variation, e.g. EST-based molecular markers (Schmitt et al. 2006); (b) functional markers (FMs) are derived from polymorphic sites within genes involved in phenotypic expression of traits, e.g. candidate gene-based molecular markers. Functional markers can further be grouped into two subgroups depending on the involvement in the phenotypic trait variation, (1) direct functional markers (DFMs), for which the role in phenotypic trait variation is well proven, and (2) indirect functional markers (IFMs), for which the role for phenotypic trait variation is indirectly known (Anderson and Lübberstedt 2003).

Few studies have been conducted on understanding the chickpea transcriptome by generating the ESTs (Boominathan et al. 2004; Romo et al. 2004; Buhariwalla et al. 2005; Coram and Pang 2005). Recently several EST sequencing projects have led to generation of large scale EST sequences through single pass sequencing (Varshney et al. 2009; Gao et al. 2008; Ashraf et al. 2009; Jain and Chattopadhyay 2010).

Several large-insert BAC libraries and binary BAC (BIBAC) libraries have been constructed in chickpea for marker development as well as construction of physical maps. For instance, 233 new chickpea SSR markers were developed by Lichtenzveig et al. (2005) by screening the BAC library with eight synthetic SSR oligos, (GA)10, (GAA)7, (AT)10, (TAA)7, (TGA)7, (CA)10, (CAA)7, and (CCA)7. Recently a set of 1,344 novel SSR markers were developed from BAC-end sequences (Thudi et al. 2011). The Chickpea Transcriptome Database (CTDB) (http://59.163.192.90:8080/ctdb/) developed at National Institute of Plant Genome Research provides user scientists/breeders a portal to search, browse and query the data to facilitate the research on chickpea and other legumes.

Quantitative Trait Loci (QTLs)

Understanding the genetics of complex traits like drought tolerance, *Helicoverpa* resistance and salinity tolerance will help in improving these traits through marker-assisted selection (MAS). Despite the importance of root traits in drought avoidance and availability of germplasm with prolific root systems such as ICC 4958 and ICC 8261, the breeding efforts to improve root traits have been negligible. This is because of the laborious, time-consuming and destructive methods involved in root studies. Molecular markers linked to major QTLs for root traits can greatly facilitate marker-assisted selection (MAS) for root traits in segregating generations. ICRISAT in collaboration with several partners generated >3,000 chickpea ESTs from a library constructed after subtractive suppressive hybridization (SSH) of root tissues from ICC 4958 and Annigeri to isolate and characterize root-specific genes differentially expressed between these genotypes (Buhariwalla et al. 2005; Jayashree et al. 2005). This database provides researchers in chickpea genomics with a major new resource for data mining associated with root traits and drought tolerance.

A set of RILs from Annigeri × ICC 4958 cross was developed at ICRISAT and characterized for root traits (Serraj et al. 2004), and SSR marker TAA 170 was identified for a major QTL that accounted for 33 % of the variation for root weight and root length (Chandra et al. 2004). Based on the screening of mini-core collection, parents genetically and phenotypically more distant were identified for development of new mapping populations. These include ICC 8261 and ICC 4958 for a large root system and ICC 283 and ICC 1882 for small root systems. These two crosses were made and more than 250 RILs were developed in each cross (Gaur et al. 2008). These two mapping populations have been phenotyped and genotyped to identify additional QTLs for root traits.

Several other intra-specific mapping populations have been developed and used to identify the markers associated with traits like resistance to fusarium wilt (Tekeoglu et al. 2002; Udupa and Baum 2003; Sharma et al. 2004, 2005), resistance to ascochyta blight (Iruela et al. 2007; Anbessa et al. 2009; Kottapalli et al. 2009), resistance to rust (Madrid et al. 2008), resistance to botrytis grey mold (Anuradha et al. 2011), salinity tolerance (Vadez et al. 2012), drought tolerance (unpublished data with ICRISAT), seed traits (Cobos et al. 2009) and, for grain yield (Rehman et al. 2012). Several of these studies have been summarized in earlier reviews (Gaur et al. 2012b; Upadhyaya et al. 2011; Varshney et al. 2007).

Genome Mapping

Physical Mapping

As mentioned above, large scale genomic resources like molecular markers and genetic linkage maps were developed during recent past. Although QTLs for different traits were identified (Table 4.2), the markers were not close enough for their effective use in molecular breeding. In this context, genome-wide physical maps have been used in several species to effectively integrate genomic tools for marker-assisted breeding, high-resolution mapping and positional cloning of genes and QTL (Chin et al. 2010). In addition physical maps will also enable desirable genome sequencing and comparative genomics. Despite these advantages, a genome-wide physical map has not been developed for chickpea. However, recently a BAC/BIBAC based physical map was developed; three large contigs closely linked to QTLs contributing to ascochyta blight resistance and flowering in chickpea were identified (Zhang et al. 2010). However, a genome-wide physical map is essential for genomics research, cloning candidate genes and enhancing molecular breeding. Towards development of genome-wide physical map, in chickpea in collaboration with National Institute of Plant Genetic Research (NIPGR), New Delhi (S Bhatia and A K Tyagi) and UC-Davis, USA (MingCheng Luo), two new BAC libraries were constructed using *Hin*dIII and *Eco*RI restriction enzymes employing pCC-1BAC Epicentre vector in DH10b. A total of 96,768 clones from both the libraries that cover ~15.7× genome were fingerprinted. In addition, clones from BAC library developed by Thudi et al. (2011) and NBS-LRR genes were also fingerprinted and used for developing the physical map as a result chickpea physical map was developed spanning an estimated 574 Mb (http://probes.pw.usda.gov:8080/chickpea/). Genetic map positions for 245 BES-SSR markers permit an initial integration of BAC contigs with the chickpea genetic map. Efforts are underway to define the minimum tiling path (MTP) based on the available physical mapping data, which will facilitate either BAC-end or pooled BAC-sequencing of MTP clones. The resulting integrated genetic and physical map is expected to enhance genetics and genomics research and breeding applications in chickpea. The integration of physical map

Table 4.2 Summary of trait mapping for biotic, abiotic and agronomically important traits in chickpea

Traits studied	QTL/genes	Markers linked	References
Biotic stress			
Resistance to *fusarium* wilt	*Foc0*	RAPD, SSR	Cobos et al. (2005)
	Foc1	SSR	Gowda et al. (2009)
	Foc2	SSR	Gowda et al. (2009)
	Foc3	SSR	Sharma et al. (2004), Gowda et al. (2009)
	Foc4	SSR	Sharma et al. (2004, 2005)
	Foc5	SSR	Gowda et al. (2009)
Ascochyta blight	QTL	RAPD	Millán et al. (2003)
	Ar19	RAPD	Rakshit et al. (2003)
	QTL*ar2b*	SSR	Udupa and Baum (2003)
	QTL*AR3*	SSR	Iruela et al. (2007)
	QTL*ar1*	SSR	Iruela et al. (2006)
	QTL*ar2*	SSR	Iruela et al. (2006)
	QTL	SSR	Anbessa et al. (2009)
	QTL	SSR	Aryamanesh et al. (2010)
Botrytis grey mould	QTL	SSR	Anuradha et al. (2011)
Resistance to rust	*Uca1/uca1*	SSR	Madrid et al. (2008)
Abiotic stress			
Salinity	QTL	SSR	Vadez et al. (2012)
Root weight; root length	QTL	SSR	Chandra et al. (2004)
Root traits	QTL	SSR	Varshney et al. (Unpublished)
Drought tolerance score	Q3-1	SSR	Rehman et al. (2012)
Canopy temperature differential	Q1-1	SSR	Rehman et al. (2012)
Agronomic and yield			
Plant growth habit	*Prostrate*	SSR	Aryamanesh et al. (2010)
	Hg/hg	RAPD	Gowda et al. (2009)
Days to flowering	Q3-1	SSR	Rehman et al. (2012)
	QTL	SSR	Aryamanesh et al. (2010)
	QTL	SSR	Aryamanesh et al. (2010)
	DF3	SSR, RAPD	Gowda et al. (2009)
Flowering time	*Efl1*, *Efl2*	–	Aryamanesh et al. (2010)
Days to maturity	Q3-1	SSR, RAPD	Rehman et al. (2012)
Seed coat thickness	QTL_{Tt}	*SSR*, *morphological*	Gowda et al. (2009)
Seed size	QTL_{SW1}	SSR	Gowda et al. (2009)
Seed/pod	*Spp*	RAPD, SSR	Radhika et al. (2007)
Double podding	*Sfl*	SSR, RAPD	Radhika et al. (2007)
Harvest index	Q1-1	SSR	Rehman et al. (2012)
	Q3-1	SSR	Rehman et al. (2012)

with genetic maps has been reported earlier in different plant species including some fruit trees such as peach (Zhebentyayeva et al. 2008), papaya (Yu et al. 2009), apple (Han et al. 2011). The framework physical map serves as a valuable resource for various other studies such as effective positional cloning of genes and quantitative trait locus (QTL) fine-mapping.

Genetic Mapping

The first linkage map of chickpea was reported in 1990 and consisted of 26 isozyme and three morphological trait loci (Gaur and Slinkard 1990a, b). Several additional isozyme loci and morphological trait loci were mapped in the subsequent studies (Simon and Muehlbauer 1997; Kazan et al. 1993; Idnani 1998). The use of DNA markers in gene mapping greatly accelerated progress in development of a detailed genetic map of chickpea. A linkage map of DNA markers was first published in 1997 which contained 10 RFLP and 45 RAPD markers (Simon and Muehlbauer 1997). These maps were developed by using F_2 mapping populations. The first map using RILs was developed in 2000, which consists of 118 STMS, 96 DAF (DNA amplification fingerprinting), 70 AFLP, 37 ISSR (inter simple sequence repeats), 17 RAPD, 2 SCAR, 3 cDNA and 8 isozyme markers (Winter et al. 2000). All these earlier studies used interspecific mapping populations because of limited polymorphism observed for then available markers in the cultivated chickpea. Availability of additional markers made it possible to use intraspecific segregations in linkage studies. A molecular map based on intraspecific cross (kabuli-desi cross) was developed and used to tag genes for resistance to Fusarium wilt. Two SCAR markers and two RAPD markers (Mayer et al. 1997) were found associated with resistance to race 1 and one ISSR marker with resistance to race 4 (Ratnaparkhe et al. 1998). The genes for resistance to races 4 and 5 were found to be linked and located close to one STMS and one SCAR marker (Winter et al. 2000).

As a result of concerted efforts of ICRISAT in collaboration with several partners across globe, large-scale markers resources are now available for chickpea. Employing these marker resources both intra and inter-specific maps have been developed. A set of interspecific RILs from *C. arietinum* (ICC 4958) × *C. reticulatum* (PI 489777) cross has been used as reference mapping population for chickpea. Nayak et al. (2010) developed a comprehensive map of this reference population with 521 loci that mainly comprised of SSR markers developed from microsatellite enriched library. Further, this map was integrated with BES-SSRs, DArT and gene-based markers by Thudi et al. (2011), which comprised of 1,291 loci. An advanced gene-rich map of chickpea comprising of 406 loci (including 177 gene-based markers) spanning 1,497.7 cM genetic distance has been developed for this reference population (Choudhary et al. 2012b). Recently, Hiremath et al. (2012) developed large-scale KASPar assays for SNP genotyping and developed a genetic map comprising 1,328 marker loci including novel 625 CKAMs (Chickpea KAspar Assay Markers), 314 TOG-SNPs and 389 published marker loci for this reference population. The summary of genetic maps developed in chickpea is illustrated in Table 4.3.

Comparative and Functional Genomics

The advances in next-generation sequencing technologies facilitated the sequencing of trancriptomes as well as the genome of several crop plants. In this context understanding the gene function is of great importance. Recently several genes/ESTs

Table 4.3 Summary of genetic maps developed for chickpea

Mapping population	No. of loci mapped	Types of markers	Genetic map length (cM)	References
ICC 4958 × PI 489777	1,328	SSR, CKAM, TOG-SNP, DArT	789	Hiremath et al. (2012)
ICC 4958 × PI 489777	406	EST-SSRs, intron targeted primers (ITPs), expressed sequence tag polymorphisms (ESTPs), and SNPs	1,498	Choudhary et al. (2012b)
ICC 4958 × PI 489777	1,063	SSR and SNP	1,809	Gaur et al. (2012a)
ICC 4958 × PI 489777	1,291	SSR, SNP, DArT	846	Thudi et al. (2011)
ICC 4958 × PI 489777	300	SSR, CISR, CAPS	767	Gujaria et al. (2011)
ICCV 2 × JG 62	138	STMS	631	Gaur et al. (2010)
ICC 4958 × PI 489777	521	SSR, RAPD, AFLP, RGA	2,602	Nayak et al. (2010)
Five narrow crosses (Desi × Kabuli types)	229	STMS, RAPD, cross-genome markers	427	Millán et al. (2010)
Five wide crosses (*C. arietinum* × *C. reticulatum*)	555	STMS, RAPD, cross-genome markers	653	Millán et al. (2010)
ICC 4991 × ICCV 04516 (F_2)	84	SSRs	724	Kottapalli et al. (2009)
JG 62 × Vijay (RIL), Vijay × ICC 4958 (RIL)	273	RAPDs and ISSRs	740	Radhika et al. (2007)
ILC72 × Cr5-10	89	RAPDs, ISSRs, STS	–	Cobos et al. (2006)
Hadas × Cr205 (RIL)	93	SSRs, CytP450 markers	345	Abbo et al. (2005)
WR315 × C104	102	ISSR, STMS, RAPD, STS	–	Sharma et al. (2004)
ILC 1272 × ILC 3279	55	SSRs	–	Udupa and Baum (2003)
ICC 12004 × Lasseter (F_2)	69	SSRs, RGAs, ISSRs	–	Flandez-Galvez et al. (2003)
Lasseter × PI 527930 (F_2)	83	RAPDs, SSRs, ISSRs, RGA	–	Collard et al. (2003)
C. arietinum × *C. reticulatum* (F_2)	296	47 defense response gene markers to the map of Winter et al. (2000)	–	Pfaff and Kahl (2003)
C. arietinum × *C. echinospermum* (F_2)	83	SSRs, RAPDs, ISSRs and RGA	–	Collard et al. (2003)
ICCV 2 × JG 62 (RIL)	103	SSRs, RAPDs, ISSRs, morphological	–	Cho et al. (2002)
ICC4958 × PI 489777 (RIL)	56	SSRs and RGA	1,175	Tekeoglu et al. (2002)
C. arietinum × *C. reticulatum* (F_2)	117	SSRs and RGA	–	Rajesh et al. (2002)
FLIP 84-92C × PI 599072 (RIL)	144	RAPDs, ISSRs, morphological, isozyme	–	Santra et al. (2000)

(continued)

Table 4.3 (continued)

Mapping population	No. of loci mapped	Types of markers	Genetic map length (cM)	References
C. arietinum × *C. reticulatum* (F_2)	116	Marker loci RAPDs, ISSRs, isozyme, and morphological	–	Santra et al. (2000)
C. arietinum × *C. reticulatum* (F_2)	354	SSRs, DAF, AFLPs, ISSRs, RAPDs, isozyme, cDNA, SCAR and morphological	2,078	Winter et al. (2000)
C. arietinum × *C. reticulatum* (F_2)	120	STMS	–	Winter et al. (1999)
C. arietinum × *C. reticulatum* (F_2); *C. arietinum* × *C. echinospermum* (F_2)	91	Morphological, isozyme, RFLPs and RAPDs	–	Simon and Muehlbauer (1997)
C. arietinum × *C. reticulatum* (F_2); *C. arietinum* × *C. echinospermum* (F_2)	28	Morphological and isozyme	–	Kazan et al. (1993)
C. arietinum × *C. reticulatum* (F_2)	29	Morphological and isozyme	–	Gaur and Slinkard (1990a), Gaur and Slinkard (1990b)

involved in various stress responses have been identified based on transcriptomic and proteomic studies (Varshney et al. 2009; Pandey et al. 2006, 2008; Mantri et al. 2007; Molina et al. 2008, 2011). However, limited efforts have been made on gene discovery and only a few candidate genes cloned and functionally validated (Kaur et al. 2008; Shukla et al. 2009; Tripathi et al. 2009; Peng et al. 2010). Several functional genomics studies have been performed in chickpea to identify the abiotic stress-responsive transcripts by approaches such as suppression subtractive hybridization (SSH), super serial analysis of gene expression (SuperSAGE), microarray, and EST sequencing (Varshney et al. 2009; Buhariwalla et al. 2005; Molina et al. 2008). The salt stress transcriptomes of roots and nodules studied by Molina et al. (2011) by using deep SuperSAGE provided deep insights into the first molecular reactions of a plant exposed to salinity. By studying two chickpea varieties (BGD 72 and ICCV 2) for differences in transcript profiling during drought stress treatment by withdrawal of irrigation at different time points. Jain and Chattopadhyay (2010) reported that most of the highly expressed ESTs in the tolerant cultivar predicted that most of them encoded proteins involved in cellular organization, protein metabolism, signal transduction, and transcription. Deokar et al. (2011) in addition to studying the genes that are up- and down-regulated in a drought-tolerant genotype (ICC 4958) under terminal drought stress and a drought susceptible genotype

(ICC 1882), also studied the gene expression between the bulks of the selected RILs exhibiting extreme phenotypes. Garg et al. (2011) reported the sequencing and *de novo* assembly of chickpea transcriptome using short-read data.

Progress Towards Whole Genome Sequencing and Data Mining

In recent years, genome sequencing has become very popular in the area of plant genomics and breeding as it offers threefold advantages: (a) enables us to understand plant genome structure and dynamics of molecular evolution, (b) enable identification of genes and functional elements and help in annotation of completed genome, and (c) provide the genomic tools and platforms for gene mapping, gene isolation and molecular breeding. Further, information gained from sequenced genomes, coupled with genetic association studies, may allow us to identify key genes/quantitative trait loci and networks in the other species. Such information can be very useful for molecular breeding programmes in order to develop improved varieties/hybrids. Several crop plant genomes have already been sequenced for instance rice (Goff et al. 2002; Yu et al. 2002), sorghum (Paterson et al. 2009), using Sanger sequencing. Further, a number of plant genomes were sequenced using NGS technologies, for example cucumber (Huang et al. 2009), castor (Chan et al. 2010), cannabis (van Bakel et al. 2011), date palm (Al-Dous et al. 2011), cocoa (Argout et al. 2011) and pigeonpea (Varshney et al. 2012b).

A draft genome sequence of chickpea has been published recently which consists of about 738-Mb draft whole genome shotgun sequence of kabuli chickpea variety CDC Frontier (Varshney et al. 2013). The sequence contains an estimated 28,269 genes. In addition, resequencing and analysis of 90 cultivated and wild genotypes from ten countries was published and targets of both breeding-associated genetic sweeps and breeding-associated balancing selection were identified. Candidate genes were identified for disease resistance and agronomic traits, including traits that distinguish desi and kabuli chickpea. The chickpea genome sequencing work was carried out by the International Chickpea Genome Sequencing Consortium (ICGSC) led by ICRISAT. This ICGSC involved 49 scientists from 23 organizations in ten countries. This is a landmark milestone in chickpea genomics and will pave the way for more rapid progress towards integrating physical and genetic maps and genomics-assisted breeding of chickpea.

Use of Genomic Resources in Molecular Breeding

The large scale genomic resources developed during recent years are currently being employed for accelerating the molecular breeding programs in chickpea. For instance, a genomic region controlling root traits and several other traits related

Table 4.4 Details of MABC progenies being developed by introgression of genomic region controlling root traits and other traits involved in drought tolerance from ICC 4958 into chickpea cultivars

Organization	Cross	Current status
EIAR, Ethiopia	Ejere × ICC 4958	BC_3F_3
	Arerti × ICC 4958	BC_3F_3
EU, Kenya	ICCV 97105 × ICC 4958	BC_3F_3
	ICCV 95423 × ICC 4958	BC_3F_4
ICRISAT, India	ICCV 10 × ICC 4958	BC_3F_4
IIPR, India	DCP92-3 × ICC 4958	BC_2F_1
	KWR108 × ICC 4958	BC_2F_1
IARI, India	Pusa 362 × ICC 4958	BC_3F_1

to drought tolerance contributing >30 % phenotypic variation identified in the Phase I of the Tropical Legume (TL-I) project of Generation Challenge Programme (GCP). The draft has been intogressed into three popular chickpea varieties, JG 11 and KAK 2 from India and Chefe from Ethiopia. Phenotypic evaluation of these lines is underway in India, Kenya and Ethiopia (Table 4.4).

A marker-assisted recurrent selection (MARS) program is also in progress at ICRISAT, India and Egerton University, Kenya for accumulating favorable alleles for yield under moisture stress conditions. MARS is a modern breeding approach that enables increasing frequency of several beneficial alleles having additive effect and small individual effects in recurrent crosses (Bernardo and Charcosset 2006). While several multi-national companies are using MARS in crops like maize and soybean, only a few public sector institutes have started to use MARS in crops likes wheat (Charmet et al. 1999), maize (Ribaut and Ragot 2007). At ICRISAT four superior desi genotypes based on their performance have been selected ICCV 04112, ICCV 05107, ICCV 93954 (released as JG 11 in India) and ICCV 94954 (released as JG 130 in India) and two crosses were made by using elite and elite lines (JG 11 × ICCV 04112 and JG 130 × ICCV 05107). The F_3 plants were genotyped and $F_{3:5}$ progenies were evaluated at three locations (Ethiopia, Kenya and India) under rainfed and irrigated conditions. To pyramid superior alleles of the favorable QTLs identified based on F_3 genotyping data and F_5 phenotyping data, a set of eight lines were selected for each cross using OptiMAS 1.0. It is anticipated that at the end of the project, RC_3F_4 progenies will be available for evaluation at multi-locations. Recently, Indian Agricultural Research Institute (IARI), New Delhi and Indian Institute of Pulses Research, Kanpur have also initiated MARS in chickpea for Pusa 372 × JG130 and DCP92-3 × ICCV 10 crosses, respectively. These efforts are expected to develop superior lines with enhanced drought tolerance.

The MAGIC population developed at ICRISAT (described in section "Mapping Populations") also provided breeding materials for direct use in chickpea breeding programs. ICRISAT has shared F_4 seed from 4-way and 8-way crosses with several institutes in South Asia and sub-Saharan Africa. The plant breeders can select promising plants at their locations and develop progenies for further evaluations. Several heat tolerant progenies have been developed from MAGIC population at ICRISAT.

Conclusions and Perspectives

Rapid advancements in development of chickpea genomics during the past decade have made it possible to initiate genomics-assisted breeding in chickpea for improvement of its adaptation to abiotic and biotic stresses. MABC lines, in which a genomic region that controls root traits and several other drought tolerance related traits was introgressed, are already under field evaluation. Several other projects on marker-assisted breeding of chickpea are in progress and elite lines being developed from these projects are expected to be available for field evaluation in coming years. The year 2013 began by adding a landmark milestone in chickpea genomics with the publication of draft genome sequence of chickpea. The information revealed by the draft genome sequence will further boost efforts on development of genomic resources and their applications in chickpea improvement. Integrated breeding approaches would improve speed, precision and efficiency of ongoing breeding efforts of chickpea improvement in development of cultivars better adapted to existing and evolving growing environments and cropping systems and with grain and nutritional quality preferred by the industry and the consumers.

References

Abbo S, Molina C, Jungmann R, Grusak MA, Berkovitch Z, Reifen R et al (2005) Quantitative trait loci governing carotenoid concentration and weight in seeds of chickpea (*Cicer arietinum* L.). Theor Appl Genet 111:185–195

Ahmad F (1999) Random amplified polymorphic DNA (RAPD) analysis reveals genetic relationships among the annual *Cicer* species. Theor Appl Genet 98:657–663

Ahmad S, Godward MBE (1980) Cytological studies on the cultivars of *Cicer arietinum* L. from Pakistan. Caryologia 33:55–68

Ahmad F, Hymowitz T (1993) The fine structure of chickpea (*Cicer arietinum* L.) chromosomes as revealed by pachytene analysis. Theor Appl Genet 86:637–641

Ahmad F, Slinkard AE (1992) Genetic relationships in the genus *Cicer* L. as revealed by polyacrylamide gel electrophoresis of seed storage proteins. Theor Appl Genet 84:688–692

Ahmad F, Gaur PM, Slinkard AE (1992) Isozyme polymorphism and phylogenetic interpretations in the genus *Cicer* L. Theor Appl Genet 83:620–627

Ahmad F, Gaur PM, Croser J (2005) Chickpea (*Cicer arietinum* L.). Genetic resources, chromosome engineering and crop improvement. In: Singh RJ, Jauhar PP (eds) Grain legumes. CRC, Boca Raton, FL, pp 187–217

Akhtar M (1954) A study on the introduction of polyploidy in gram. Proc Pak Sci Conf 6:124–125

Al-Dous EK, George B, Al-Mahmoud ME, Al-Jaber MY, Wang H, Salameh YM et al (2011) De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). Nat Biotechnol 29:521–527

Anbessa Y, Taran B, Warkentin TD, Tullu A, Vandenberg A (2009) Genetic analyses and conservation of QTL for ascochyta blight resistance in chickpea. Theor Appl Genet 119:757–765

Anderson JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–560

Anuradha C, Gaur PM, Pande S, Gali KK, Ganesh M, Kumar J et al (2011) Mapping QTL for resistance to botrytis grey mould in chickpea. Euphytica 182:1–9

Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J et al (2011) The genome of Theobroma ca-cao. Nat Genet 43:101–108

Aryamanesh N, Nelson MN, Yan G, Clarke HJ, Siddique KHM (2010) Mapping a major gene for growth habit and QTLs for ascochyta blight resistance and flowering time in a population between chickpea and *Cicer reticulatum*. Euphytica 173:307–319

Ashraf N, Ghai D, Barman P, Basu S, Gangisetty N, Mandal MK, Chakraborty N, Datta A, Chakraborty S (2009) Comparative analyses of genotype dependent expressed sequence tags and stress-responsive transcriptome of chickpea wilt illustrate predicted and unexpected genes and novel regulators of plant immunity. BMC Genomics 10:415

Bernardo R, Charcosset A (2006) Usefulness of gene information in marker-assisted recurrent selection: a simulation appraisal. Crop Sci 46:614–621

Boominathan P, Shukla R, Kumar A, Manna D, Negi D, Verma PK, Chattopadhyay D (2004) Long term transcript accumulation during the development of dehydration adaptation in *Cicer arietinum*. Plant Physiol 135:1608–1620

Buhariwalla HK, Jayashree B, Eshwar K, Crouch JH (2005) Development of ESTs from chickpea roots and their use in diversity analysis of the *Cicer* genus. BMC Plant Biol 5:16

Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D et al (2010) Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotechnol 28:951–956

Chandra S, Buhariwalla HK, Kashiwagi J, Harikrishna S, Sridevi KR, Krishnamurthy L et al (2004) Identifying QTL-linked markers in marker-deficient crops. In: 4th International crop science congress, Brisbane, Australia, 26 Sep–1 Oct 2004

Charmet G, Robert N, Perretant MR, Gay G, Sourdille P, Groos C et al (1999) Marker-assisted recurrent selection for cumulating additive and interactive QTLs in recombinant inbred lines. Theor Appl Genet 99:1143–1148

Chin JH, Lu X, Haefele SM, Gamuyao R, Ismail A et al (2010) Development and application of gene-based markers for the major rice QTL Phosphorus uptake 1. Theor Appl Genet 120:1073–1086

Cho S, Kumar J, Shultz JF, Anupama K, Tefera F, Muehlbauer FJ (2002) Mapping genes for double podding and other morphological traits in chickpea. Euphytica 125:285–292

Choudhary S, Sethy NK, Shokeen B, Bhatia S (2006) Development of sequence-tagged microsatellites site markers for chickpea (*Cicer arietinum* L.). Mol Ecol Notes 6:93–95

Choudhary P, Khanna SM, Jain PK, Bharadwaj C, Kumar J, Lakhera PC et al (2012a) Genetic structure and diversity analysis of the primary gene pool of chickpea using SSR markers. Genet Mol Res 11:891–905

Choudhary S, Gaur R, Gupta S, Bhatia S (2012b) EST-derived genic molecular markers: development and utilization for generating an advanced transcript map of chickpea. Theor Appl Genet 124:1449–1462

Cobos MJ, Fernandez MJ, Rubio J, Kharrat M, Moreno MT, Gil J et al (2005) A linkage map in chickpea (*Cicer arietinum* L.) in two populations from Kabuli × Desi crosses: location of a resistance gene for fusarium wilt race 0. Theor Appl Genet 110:1347–1353

Cobos MJ, Rubio J, Strange RN, Moreno MT, Gil J, Millan T (2006) A new QTL for Ascochyta blight resistance in an RIL population derived from an interspecific cross in chickpea. Euphytica 149:105–111

Cobos MJ, Winter P, Kharrat M, Cubero JI, Gil J, Millan T et al (2009) Genetic analysis of agronomic traits in a wide cross of chickpea. Field Crops Res 111:130–136

Collard BCY, Pang ECK, Ades PK, Taylor PWJ (2003) Preliminary investigations of QTL associated with seedlings resistance to Ascochyta blight from *Cicer echinospermum*, a wild relative of chick pea. Theor Appl Genet 107:719–729

Coram T, Pang E (2005) Isolation and analysis of candidate ascochyta blight defense genes in chickpea, Part I. Generation and analysis of an expressed sequence tag (EST) library. Physiol Mol Plant Pathol 66:192–200

Croser JS, Ahmad F, Clarke HJ, Siddique KHM (2003) Utilization of wild *Cicer* in chickpea improvement—progress, constraints and prospects. Aus J Agric Res 54:429–444

Deokar AA, Kondawar V, Jain PK, Karuppayil SM, Raju NL, Vadez V et al (2011) Comparative analysis of expressed sequence tags (ESTs) between drought-tolerant and -susceptible genotypes of chickpea under terminal drought stress. BMC Plant Biol 11:70

FAOSTAT.database(2012)http://faostat.fao.org/site/567/DextopDEfault.aspx?PageID=567#ancor. Accessed 14 Mar 2013

Flandez-Galvez H, Ades PK, Ford R, Pang ECK, Taylor PWJ (2003) QTL analysis for ascochyta blight resistance in an intraspecific population of chickpea (*Cicer arietinum* L.). Theor Appl Genet 107:1257–1265

Gao WR, Wang XS, Liu QY, Peng H, Chen C, Li JG, Zhang JS, Hu SN, Ma H (2008) Comparative analysis of ESTs in response to drought stress in chickpea (*C. arietinum* L.). Biochem Biophys Res Commun 376:578–583

Garg R, Patel RK, Tyagi AK, Jain M (2011) *De Novo* assembly of chickpea transcriptome using short reads for gene discovery and marker identification. DNA Res 18:53–63

Gaur PM, Slinkard AE (1990a) Inheritance and linkage of isozyme coding genes in chickpea. J Hered 81:455–461

Gaur PM, Slinkard AE (1990b) Genetic control and linkage relations of additional isozyme markers in chickpea. Theor Appl Genet 80:648–656

Gaur PM, Krishnamurthy M, Kashiwagi J (2008) Improving drought-avoidance root traits in chickpea (*C. arietinum*)—current status of research at ICRISAT. Plant Prod Sci 11:3–11

Gaur PM, Mallikarjuna N, Knights T, Beebe S, Debouck D, Mejía A et al (2010) Gene introgression in grain legumes. In: Gupta S, Ali M, Singh BB (eds) Grain legumes: genetic improvement, management and trade. Indian Society of Pulses Research and Development, Indian Institute of Pulses Research, Kanpur, India, pp 1–17

Gaur PM, Jukanti AK, Varshney RK (2012) Impact of genomic technologies on chickpea breeding strategies. Agronomy 2:199–221

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. Japonica). Science 296:92–100

Gowda SJM, Radhika P, Kadoo NY, Mhase LB, Gupta VS (2009) Molecular mapping of wilt resistance genes in chickpea. Mol Breed 24:177–183

Gujaria N, Kumar A, Dauthal P, Dubey A, Hiremath P, Bhanu Prakash A et al (2011) Development and use of genic molecular markers (GMMs) for construction of a transcript map of chickpea (*Cicer arietinum* L.). Theor Appl Genet 122:1577–1589

Gupta PK, Bahl JR (1983) Cytogenetics and origin of some pulse crops. In: Swaminathan MS, Gupta PK, Sinha U (eds) Cytogenetics of crop plants. Macmillan India Ltd, New Delhi, India, pp 405–440

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163–185

Han Y, Zheng D, Vimolmangkang S, Khan MA, Beever JE, Korban SS (2011) Integration of physical and genetic maps in apple confirms whole-genome and segmental duplications in the apple genome. J Exp Bot 62:5117–5130

Hiremath PJ, Kumar A, Penmetsa RV, Farmer A, Schlueter JA et al (2012) Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. Plant Biotechnol J 10:716–732

Huang S, Li R, Zhang Z, Li L, Gu XF, Fan W et al (2009) The genome of the cucumber, *Cucumis sativus* L. Nat Genet 41:1275–1281

Huttel B, Winter P, Weising K, Choumane W, Weigand F, Kahl G (1999) Sequence-tagged microsatellite site markers for chickpea (*Cicer arietinum* L.). Genome 42:210–217

Idnani N (1998) Inheritance and linkage studies with some isozyme and morphological loci in chickpea. M.Sc. Dissertation, JNKVV, India

Iruela M, Rubio J, Cubero JI, Gil J, Milan T (2002) Phylogenetic analysis in the genus *Cicer* and cultivated chickpea using RAPD and ISSR markers. Theor Appl Genet 104:643–651

Iruela M, Rubio J, Barro F, Cubero JI, Millan T, Gil J (2006) Detection of two quantitative trait loci for resistance to Ascochyta blight in an intra-specific cross of chickpea (*Cicer arietinum* L.): development of SCAR markers associated with resistance. Theor Appl Genet 112:278–287

Iruela M, Castro P, Rubio J, Cubero JI, Jacinto C, Millan T et al (2007) Validation of a QTL for resistance to ascochyta blight linked to resistance to fusarium wilt race 5 in chickpea (*Cicer arietinum* L.). Eur J Plant Pathol 119:29–37

Jain D, Chattopadhyay D (2010) Analysis of gene expression in response to water deficit of chickpea (*Cicer arietinum* L.) varieties differing in drought tolerance. BMC Plant Biol 10:24

Jayashree B, Buhariwalla HK, Shinde S, Crouch JH (2005) A legume genomics resource: the chickpea root expressed sequence tag database. Electron J Biotechnol 8. http://www.ejbiotechnology.info/content/vol2/issue3/full/3/index.html. Accessed 22 Mar 2013

Kaur H, Shukla RK, Yadav G, Chattopadhyay D, Majee M (2008) Two divergent genes encoding L-myo-inositol 1-phosphate synthase1 (CaMIPS1) and 2 (CaMIPS2) are differentially expressed in chickpea. Plant Cell Environ 31:1701–1716

Kazan K, Muehlbauer FJ (1991) Allozyme variation and phylogeny in annual species of *Cicer* (Leguminosae). Plant Syst Evol 175:11–21

Kazan K, Muehlbauer FJ, Weeden NF, Ladizinsky G (1993) Inheritance and linkage relationships of morphological and isozyme loci in chickpea (*Cicer arietinum* L.). Theor Appl Genet 86:417–426

Kottapalli P, Gaur PM, Katiyar SK, Crouch JH, Buhariwalla HK, Pande S et al (2009) Mapping and validation of QTLs for resistance to an Indian isolate of ascochyta blight pathogen in chickpea. Euphytica 165:79–88

Labdi M, Robertson LD, Singh KB, Charrier A (1996) Genetic diversity and phylogenetic relationships among the annual *Cicer* species as revealed by isozyme polymorphism. Euphytica 88:181–188

Lichtenzveig J, Scheuring C, Dodge J, Abbo S, Zhang HB (2005) Construction of BAC and BIBAC libraries and their applications for generation of SSR markers for genome analysis of chickpea, *Cicer arietinum* L. Theor Appl Genet 110:492–510

Madrid E, Rubiales D, Moral A, Moreno MT, Millan T, Gil J et al (2008) Mechanism and molecular markers associated with rust resistance in a chickpea interspecific cross (*Cicer arietinum* × *Cicer reticulatum*). Eur J Plant Pathol 121:43–53

Mantri NL, Ford R, Coram TE, Pang EC (2007) Transcriptional profiling of chickpea genes differentially regulated in response to high-salinity, cold and drought. BMC Genomics 8:303

Mayer MS, Tullu A, Simon CJ, Kumar J, Kaiser WJ, Kraft JM et al (1997) Development of a DNA marker for Fusarium wilt resistance in chickpea. Crop Sci 37:1625–1629

Millán T, Rubio J, Iruela M, Daly K, Cubero JI, Gil J (2003) Markers associated with Ascochyta blight resistance in chickpea and their potential in marker-assisted selection. Field Crops Res 84:373–384

Millan T, Clarke HJ, Siddique KHM, Buhariwalla HK, Gaur PM et al (2006) Chickpea molecular breeding: new tools and concepts. Euphytica 147:81–103

Millán T, Winter P, Jungling R, Gil J, Rubio J, Cho S et al (2010) A consensus genetic map of chickpea (*Cicer arietinum* L.) based on 10 mapping populations. Euphytica 175:175–189

Molina C, Rotter B, Horres R, Udupa SM, Besser B, Bellarmino L et al (2008) Super SAGE: the drought stress-responsive transcriptome of chickpea roots. BMC Genomics 9:553

Molina C, Zaman-Allah M, Khan F, Fatnassi N, Horres R, Rotter B et al (2011) The salt-responsive transcriptome of chickpea roots and nodules via deepSuperSAGE. BMC Plant Biol 11:31

Nayak SN, Zhu H, Varghese N, Datta S, Choi HK, Horres R et al (2010) Integration of novel SSR and gene-based SNP marker loci in the chickpea genetic map and establishment of new anchor points with *Medicago truncatula* genome. Theor Appl Genet 120:1415–1441

Nguyen TT, Taylor PWJ, Redden RJ, Ford R (2004) Genetic diversity estimates in *Cicer* using AFLP analysis. Plant Breed 123:173–179

Palomino C, Fernández-Romero MD, Rubio J, Torres A, Moreno MT, Millán T (2009) Integration of new CAPS and dCAPS-RGA markers into a composite chickpea genetic map and their association with disease resistance. Theor Appl Genet 118:671–682

Pandey A, Choudhary MK, Bhushan D, Chattopadhyay A, Chakraborty S, Datta A, Chakraborty N (2006) The nuclear proteome of chickpea (*Cicer arietinum* L.) reveals predicted and unexpected proteins. J Proteome Res 5:3301–3311

Pandey A, Chakraborty S, Datta A, Chakraborty N (2008) Proteomics approach to identify dehydration responsive nuclear proteins from chickpea (*Cicer arietinum* L.). Mol Cell Proteomics 7:88–107

Parsons BL, Heflich RH (1997) Genotypic selection methods for the direct analysis of point mutations. Mutat Res 387:97–121

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Peng H, Yu X, Cheng H, Shi Q, Zhang H, Li J et al (2010) Cloning and characterization of a novel NAC family gene CarNAC1 from chickpea (*Cicer arietinum* L.). Mol Biotechnol 44:30–40

Pfaff T, Kahl G (2003) Mapping of gene-specific markers on the genetic map of chickpea (*Cicer arietinum* L.). Mol Genet Genomics 269:243–251

Pundir RPS, Mengesha MH (1983) Collection of chickpea germplasm in Ethiopia. Int Chickpea Newsl 8:6–7

Radhika P, Gowda S, Kadoo N, Mhase L, Jamadagni B, Sainani M et al (2007) Development of an integrated intraspecific map of chickpea (*Cicer arietinum* L) using two recombinant inbred line populations. Theor Appl Genet 115:209–216

Rajesh PN, Tullu A, Gil J, Gupta VS, Ranjekar PK, Muehlbauer FJ (2002) Identification of an STMS marker for the doublepodding gene in chickpea. Theor Appl Genet 105:604–607

Rajesh PN, Bleness MO, Roe BA, Muehlbauer FJ (2008) Analysis of genome organization, composition and microsynteny using 500 kb BAC sequences in chickpea. Theor Appl Genet 117:449–458

Rakshit S, Winter P, Tekeoglu M, Juarez Muñoz J, Pfaff T, Benko-Iseppon AM et al (2003) DAF marker tightly linked to a major locus for ascochyta blight resistance in chickpea (*Cicer arietinum* L.). Euphytica 132:23–30

Ramanujam S, Joshi AB (1941) Colchicine-induced polyploidy in crop plants. Gram (*Cicer arietinum* L.). Indian J Agric Sci 11:835–849

Rao L, Usha Rani P, Deshmukh P, Kumar P, Panguluri S (2006) RAPD and ISSR fingerprinting in cultivated chickpea (*Cicer arietinum* L.) and its wild progenitor *Cicer reticulatum* Ladizinsky. Genet Resour Crop Evol 54:1235–1244

Ratnaparkhe MB, Santra DK, Tullu A, Muehlbauer FJ (1998) Inheritance of inter-simple-sequence-repeat polymorphisms and linkage with a fusarium wilt resistance gene in chickpea. Theor Appl Genet 96:348–353

Rehman AU, Malhotra RS, Bett K, Tar'an B, Bueckert R, Warkentin TD (2012) Mapping QTL associated with traits affecting grain yield in chickpea (*Cicer arietinum* L.) under terminal drought stress. Crop Sci 51:450–463

Ribaut JM, Ragot M (2007) Marker-assisted selection to improve drought adaptation in maize: the backcross approach, perspectives, limitations, and alternatives. J Exp Bot 58:351–360

Romo S, Labrador E, Dopico B (2004) Water stress-regulated gene expression in *Cicer arietinum* seedlings and plants. J Plant Physiol Biochem 39:1017–1026

Santra DK, Tekeoglu M, Ratnaparkhe M, Kaiser WJ, Muehlbauer FJ (2000) Identification and mapping of QTLs conferring resistance to ascochyta blight in chickpea. Crop Sci 40: 1606–1612

Schmitt BA, Costa JH, de Melo DF (2006) AOX—a functional marker for efficient cell reprogramming under stress. Trends Plant Sci 11:281–287

Sen NK, Jana MV (1956) A spontaneous autotetraploid gram (*Cicer arietinum* L.). Curr Sci 25:231–252

Serraj R, Krishnamurthy L, Kashiwagi J, Kumar J, Chandra S, Crouch JH (2004) Variation in root traits of chickpea (*Cicer arietinum* L.) grown under terminal drought. Field Crops Res 88:115–127

Serret MD, Udupa SM, Weigand F (1997) Assessment of genetic diversity of cultivated chickpea using microsatellite derived RFLP markers: implications for origin. Plant Breed 116:573–578

Sharma PC, Gupta PK (1982) Chromosome associations in autotetraploid chickpea. Int Chickpea Newsl 6:3–4

Sharma PC, Gupta PK (1983) Cytological studies in the genus *Cicer* L. Proc of XV Inter Cong Genet, 12–21 Dec 1983, Oxford & IBH Pub Co, New Delhi, India. Abstract No. 1257

Sharma PC, Winter P, Bünger T, Hüttel B, Weigand F, Weising K et al (1995) Abundance and polymorphism of di-, tri-, and tetra-nucleotide tandem repeats in chickpea (*Cicer arietinum* L.). Theor Appl Genet 90:90–96

Sharma KD, Winter P, Kahl G, Muehlbauer FJ (2004) Molecular mapping of *Fusarium oxysporum* f. sp. *ciceris* race 3 resistance gene in chickpea. Theor Appl Genet 108:1243–1248

Sharma KD, Chen W, Muehlbauer FJ (2005) Genetics of chickpea resistance to five races of Fusarium wilt and a concise set of race differentials for *Fusarium oxysporum* f. sp. *ciceris*. Plant Dis 89:385–390

Shukla RK, Tripathi V, Jain D, Yadav RK, Chattopadhyay D (2009) CAP2 enhances germination of transgenic tobacco seeds at high temperature and promotes heat stress tolerance in yeast. FEBS J 276:5252–5262

Simon CJ, Muehlbauer FJ (1997) Construction of a chickpea linkage map and its comparison with maps of pea and lentil. J Hered 88:115–119

Sudupak MA (2004) Inter- and intra-species inter simple sequence repeat (ISSR) variation in the genus *Cicer*. Euphytica 135:229–238

Sudupak MA, Akkaya MS, Kence A (2002) Analysis of genetic relationships among perennial and annual *Cicer* species growing in Turkey using RAPD markers. Theor Appl Genet 105:1220–1228

Talebi R, Naji AM, Fayaz F (2008) Geographical patterns of genetic diversity in cultivated chickpea (*Cicer arietinum* L.) characterized by amplified fragment length polymorphism. Plant Soil Environ 54:447–452

Talebi R, Jelodar NB, Fayyaz F, Mardi M (2009) AFLP-based fingerprinting of diverse chickpea genotypes: origin, history of chickpea. New Biotechnol 25:S302

Tayyar RI, Waines JG (1996) Genetic relationships among annual species of *Cicer* (Fabaceae) using isozyme markers. Theor Appl Genet 92:245–254

Tekeoglu M, Rajesh PN, Muehlbauer FJ (2002) Integration of sequence tagged microsatellite sites to the chickpea genetic map. Theor Appl Genet 105:847–854

Thudi M, Bohra A, Nayak SN, Varghese N, Shah TM, Penmetsa RV et al (2011) Novel SSR markers from BAC-end sequences, DArT arrays and a comprehensive genetic map with 1,291 marker loci for chickpea (*Cicer arietinum* L.). PLoS One 6:e27275

Thudi M, Li Y, Jackson SA, May GD, Varshney RK (2012) Current state-of-art of sequencing technologies for plant genomics research. Brief Funct Genomics 11:3–11

Tripathi V, Parasuraman B, Laxmi A, Chattopadhyay D (2009) CIPK6, a CBL-interacting protein kinase is required for development and salt tolerance in plants. Plant J 58:778–790

Udupa SM, Baum M (2003) Genetic dissection of pathotype-specific resistance to ascochyta blight disease in chickpea (*Cicer arietinum* L.) using microsatellite markers. Theor Appl Genet 106:1196–1202

Udupa SM, Sharma A, Sharma RP, Pai RA (1993) Narrow genetic variability in *Cicer arietinum* L. as revealed by RFLP analysis. J Plant Biochem Biotechnol 2:83–86

Upadhyaya HD, Thudi M, Dronavalli N, Gujaria N, Singh S, Sharma S et al (2011) Genomic tools and germplasm diversity for chickpea improvement. Plant Genet Resour 9:45–48

Vadez V, Krishnamurthy L, Thudi M, Anuradha C, Colmer TD, Turner NC et al (2012) Assessment of ICCV 2×JG 62 chickpea progenies shows sensitivity of reproduction to salt stress and reveals QTL for seed yield and yield components. Mol Breed 30:9–21

van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR et al (2011) The draft genome and transcriptome of Cannabis sativa. Genome Biol 12:R102

Van der Maesen LJG, Pundir RPS (1987) Availability and use of wild *Cicer* germplasm. Plant Genet Resour Newsl 57:19–24

Varshney RK, Hoisington DA, Upadhyaya HD, Gaur PM, Nigam SN, Saxena K et al (2007) Molecular genetics and breeding of grain legume crops for the semi-arid tropics. In: Varshney RK, Tuberosa R (eds) Genomics-assisted crop improvement, vol II, Genomics applications in crops. Springer, The Netherlands, pp 207–242

Varshney RK, Hiremath PJ, Lekha P, Kashiwagi J, Balaji J, Deokar AA et al (2009) A comprehensive resource of drought-and salinity-responsive ESTs for gene discovery and marker development in chickpea (*Cicer arietinum* L). BMC Genomics 10:523

Varshney RK, Thudi M, May GD, Jackson SA (2010) Legume genomics and breeding. Plant Breed Rev 33:257–304

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA et al (2012a) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource poor farmers. Nat Biotechnol 30:83–89

Varshney RK, Kudapa H, Roorkiwal M, Thudi M, Pandey MK, Saxena RK (2012b) Advances in genetics and molecular breeding of three legume crops of semi-arid tropics using next-generation sequencing and high-throughput genotyping technologies. J Biosci 37:1–10

Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG et al (2013) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. Nat Biotechnol 31:240–246

Weber JL (1990) The informativeness of human (dC-dA)n, (dGdT)n polymorphisms. Genomics 7:524–539

Weiland JJ, Yu MH (2003) A cleaved amplified polymorphic sequence (CAPS) marker associated with root-knot nematode resistance in sugarbeet. Crop Sci 43:1814–1818

Winter P, Pfaff T, Udupa SM, Huttel B, Sharma PC et al (1999) Characterization and mapping of sequence-tagged Microsatellite sites in the chickpea (*Cicer arietinum* L.) genome. Mol Gen Genet 262:90–101

Winter P, Benko-Iseppon AM, Hettel B, Ratnaparkhe M, Tullu A et al (2000) A linkage map of the chickpea (*Cicer arietinum* L.) genome based on recombinant inbred lines from a C. *arietinum*, C. *reticulatum* cross: localization of resistance genes for fusarium wilt races 4 and 5. Theor Appl Genet 101:1155–1163

Yu J, Hu S, Wang J, Wong GKS, Li S, Liu B et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296:79–92

Yu Q, Tong E, Skelton RL, Bowers JE, Jones MR, Murray JE et al (2009) A physical map of the papaya genome with integrated genetic map and genome sequence. BMC Genomics 10:371

Zhang X, Scheuring CF, Zhang M, Dong JJ, Zhang Y, Huang JJ et al (2010) A BAC/BIBAC-based physical map of chickpea (*Cicer arietinum* L.). BMC Genomics 11:501

Zhebentyayeva TN, Swire-Clark G, Georgi LL, Garay L, Jung S, Forrest S et al (2008) A framework physical map for peach, a model Rosaceae species. Tree Genet Genomes 4:745–756

Chapter 5
Advances in Pigeonpea Genomics

Abhishek Bohra, Rachit K. Saxena, K.B. Saxena, C.V. Sameerkumar, and Rajeev K. Varshney

Abstract Pigeonpea, a member of family *Fabaceae*, is one of the important food legumes cultivated in tropical and subtropical regions. Due to its inherent properties to withstand harsh environments, it plays a critical role in ensuring sustainability in the subsistence agriculture. Furthermore, plasticity in the maturity duration imparts it a greater adaptability in a variety of cropping systems. In the post genomics era, the importance of pigeonpea is further evident from the fact that pigeonpea has emerged as first non-industrial legume crop for which the whole genome sequence has been completed. It revealed 605.78 Mb of assembled and anchored sequence as against the predicted 833 Mb genome consequently representing 72.8 % of the whole genome. In order to perform genetic and genomic analysis various molecular markers like random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR), diversity array technology (DArT), single feature polymorphism (SFP), and single nucleotide polymorphism (SNP) were employed. So far four transcriptome assemblies have been constructed and different sets of EST-SSRs were developed and validated in a panel of diverse pigeonpea genotypes. Extensive survey of BAC-end sequences (BESs) provided 3,072 BES-SSRs and all these BES-SSRs were further used for linkage analysis and trait mapping. To make the available linkage information more useful, six intra-specific genetic maps were joined together into a single consensus genetic map providing map positions to a total of 339 SSR markers

A. Bohra, Ph.D.
Crop Improvement Division, Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

R.K. Saxena, Ph.D. • R.K. Varshney, Ph.D. (✉)
Center of Excellence in Genomics (CEG), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Hyderabad, Andhra Pradesh 502324, India
e-mail: r.k.varshney@cgiar.org

K.B. Saxena, Ph.D. • C.V. Sameerkumar, Ph.D.
Pigeonpea Breeding, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, Andhra Pradesh, India

S. Gupta et al. (eds.), *Legumes in the Omic Era*, DOI 10.1007/978-1-4614-8370-0_5,

at map coverage of 1,059 cM. However, earlier very few linkage maps were available in the crop because of non-availability of genomic resources. Several quantitative trait loci (QTLs) associated with traits of agronomic interest including QTLs for sterility mosaic disease, fertility restoration, plant type and earliness have been identified and validated. To strengthen the traditional breeding, plenty of genomics tools and technologies are now available for integration in regular pigeonpea breeding schemes. This article presents the progress made in the area of pigeonpea genomics and outlines its applications in crop breeding for pigeonpea improvement.

Keywords Pigeonpea • Genetic map • Quantitative trait loci • Marker assisted selection • Genome sequence

Introduction

Pigeonpea [*Cajanus cajan* (L.) Millsp] is one of the most important legume species belonging to the family *Papilionoideae* and a member of warm season legumes (Millettioid clade). Pigeonpea is grown mainly in Asia, Africa and Central/South America, on ~5 million hectares. India ranks first in pigeonpea production with 2.46 million metric tons (mmt) followed by Myanmar (0.2 mmt) and Malawi (0.18 mmt) (FAOSTAT 2011). In developing world specially in India and Africa, pigeonpea remains one of the potential sources for livelihood generation and providing proteins to the resource poor farmers, whereas, in other countries such as Myanmar and China, it is gaining importance as one of the commodity crop to generate the foreign revenue. The cultivation of pigeonpea mostly in marginal and degraded soils and risk prone environments often causes considerable reduction in crop yield due to several factors. These factors mainly include diseases, insects/pests and abiotic stresses such as drought, salinity and water logging. This has reflected in form of a wide yield gap existing between the potential yield and actual yield realized at farmers' field (see Varshney et al. 2012).

Realizing its importance in subsistence agriculture, sincere efforts have been directed towards genetic improvement of pigeonpea. Significant genetic gains have been achieved in the form of release of several pureline varieties along with cytoplasmic genetic male-sterility (CGMS) based hybrids that has led to the expansion of production area from 2.7 mha (1961) to 5.83 mha (2011) however average yield still remains in the range of 736–755 kg/ha (FAOSTAT 2011). Domestication and breeding methods focusing strictly on self-pollination led to drastic narrowing down of the genetic base therefore further complicated the situation.

In order to experience a quantum jump in the productivity, traditional breeding efforts should be supplemented with the genomics technologies. All the essential prerequisites such as large scale DNA markers e.g. simple sequence repeat markers (SSRs), diversity array technology markers (DArT), single feature polymorphisms (SFPs) and single nucleotide polymorphisms (SNPs), genetic and quantitative trait loci (QTLs) maps, trait specific mapping populations and sequence information (transcriptome and genome assemblies) are now available in pigeonpea for

Table 5.1 Genomic resources in pigeonpea

Genomic resources	Number	Features	References
Mapping populations	~30	Segregating for *Fusarium* wilt (FW), sterility mosaic disease (SMD), water logging tolerance and plant type	Kumawat et al. (2012), Varshney et al. (2010a), Dhanasekar et al. (2010), Kotresh et al. (2006)
BAC resources			
1) BAC libraries	2	Comprising 34,560 clones each with 11× coverage of pigeonpea genome	Bohra et al. (2011)
2) BAC-end sequences	88,860	A set of >52K non-redundant sequences represented 35 Mb or ~4.3 % of the pigeonpea genome	Bohra et al. (2011)
Second and third generation DNA markers			
1) SSRs			
a) Genomic (gSSRs)	~3,300	BAC library and BES-derived highly informative SSRs	Odeny et al. (2007), (2009), Saxena et al. (2010b), Bohra et al. (2011)
b) Genic or EST-SSRs			
i) Sanger sequencing	84	Average polymorphic information content (PIC) value of 0.40	Raju et al. (2010)
ii) Deep transcriptome sequencing	550	PIC values ranged from 0.46 to 0.72	Dutta et al. (2011)
c) *In silico* mining of draft genome sequence	23,410	Containing tri-, tetra-, penta-, hexa- or compound repeat units	Varshney et al. (2012)
2) SNPs	28,104	Specific to parental combinations derived from 12 genotypes	Varshney et al. (2012)
3) DArTs	29,000	Diversity surveyed for 400 genotypes	Varshney et al. (2010b)
4) SFPs	5,692	Specific for drought tolerance	Saxena et al. (2011)
Genetic maps			
1) DArTs based	1 paternal and 1 maternal	Maps covered 270.0 cM and 451.6 cM of the total genome	Yang et al. (2009)
2) SSRs based	7 (1 inter-specific and 6 intra-specific)	Covering map distances from 466.97 to 930.9 cM	Bohra et al. (2011), (2012), Gnanesh et al. (2011)

(continued)

Table 5.1 (continued)

Genomic resources	Number	Features	References
3) SNPs based	1 intra-specific	Total map length 1520.22 cM	Kumawat et al. (2012)
	1 inter-specific	Total map length 996.21 cM	Saxena et al. (2012)
Transcriptomic resources			
1) ESTs	25,314	Sanger as well as third generation sequencing derived	http://www.ncbi.nlm.nih.gov
2) Transcriptome assemblies	4	Number of transcript assembly contigs (TACs) ranging from 4,557 to 48,726	Kudapa et al. (2012)
Whole genome sequence			
Draft genome sequences of variety 'Asha'	2	~605.78 Mb of genome with ~163.4× coverage	Varshney et al. (2012)
		~511 Mb of the genome with ~10× coverage	Singh et al. (2011)

initiating genomics assisted breeding (GAB) (Bohra et al. 2011; Varshney et al. 2012; Saxena et al. 2012) (Table 5.1). In recent years, several novel molecular breeding methodologies have been proposed for the crop improvement such as marker assisted back-crossing (MABC) and marker assisted recurrent selection (MARS) which offer a precise manner to choose a desired/superior genotype (Varshney et al. 2013). Approaches like multi parent advance generation inter-cross (MAGIC) and introgression libraries (ILs) are offering new avenues to tap natural genetic variation available in wild relatives and landraces into the cultivated gene pool (Varshney et al. 2013).

This chapter provides an overview on availability of genomic resources and the current status of molecular breeding approaches in pigeonpea improvement and explores possibilities to implement emerging molecular genetics and breeding approaches to gain the advancement in pigeonpea research and productivity.

Genome Size

Pigeonpea is a diploid crop with chromosome number $2n = 2x = 22$. The various karyotype studies conducted in pigeonpea (Krishnaswamy and Ayyangar 1935; Naithani 1941; Akinola et al. 1972) have concluded that all the wild relatives of pigeonpea carry the same number of chromosomes. After soybean, pigeonpea became the second member of clade *Phaseoloid* for which the draft genome sequence has become available and based on K-mer statistics the entire genome size was estimated to be 833.07 Mb (Varshney et al. 2012).

Genomic Resources

Mapping Populations

Availability of large segregating populations is an essential requirement for molecular tagging of traits of interest. Several types of bi-parental mapping populations such as F_2, Backcross (BC_1F_1), recombinant inbred lines (RILs), near isogenic lines (NILs) and double haploid (DH) are being employed for genetic map construction and trait mapping. Based on morphological and molecular diversity and targeting the trait segregation a series of mapping populations were generated in pigeonpea under phase I of pigeonpea genomic initiative (PGI). A total of 25 F_2 mapping populations were reported in pigeonpea segregating for several traits such as resistance to sterility mosaic disease (SMD), *Fusarium* wilt (FW), water logging and fertility restoration (Rf). Most of these populations have reached to the RILs and are being deployed for multi-location trials. Details on these mapping populations have been provided by Varshney et al. (2010a). Of these mapping populations, an inter-specific F_2 mapping population (ICP 28 × ICPW 94) was chosen for constructing high density reference genetic map for pigeonpea (Bohra et al. 2011; Saxena et al. 2012). Apart from PGI, few more mapping populations were developed at various national agricultural research centers (Kotresh et al. 2006; Dhanasekar et al. 2010; Ganapathy et al. 2009; Kumawat et al. 2012) (Table 5.2).

Molecular Markers

A wide range of DNA markers have been employed in pigeonpea including RAPD (Ratnaparkhe et al. 1995), RFLP (Sivaramakrishnan et al. 1997, 2002), AFLP (Panguluri et al. 2005), SSR (Saxena et al. 2010a; Bohra et al. 2011), DArT (Yang et al. 2006, 2011), SFP (Saxena et al. 2011) and SNP (Varshney et al. 2012; Saxena et al. 2012) etc. All these marker systems have been used for a variety of applications e.g. estimation of genetic diversity, construction of genetic maps, etc. in pigeonpea. Initially SSRs were preferred over other marker systems due to unavailability of SNPs and several advantages like higher abundance, co-dominant and multi-allelic nature and ease of scoring etc. In pigeonpea, SSRs were generated through (1) enriched library (Burns et al. 2001; Saxena et al. 2010a) (2) *in silico* expressed sequence tags (ESTs) mining (Dutta et al. 2011; Dubey et al. 2011) and (3) surveying BAC-end sequences and whole genome sequence (Bohra et al. 2011; Varshney et al. 2012). The first set of SSRs comprising ten SSRs in pigeonpea was developed by Burns et al. (2001) using CA and CT repeat enriched libraries. However, development of SSRs through enriched libraries remains to be time consuming and of low through put. In this context, sequencing of BAC ends and mining for SSRs had provided potential alternative for large scale SSR discovery.

Table 5.2 Trait mapping in pigeonpea

Name of population	Type of population	Size of population	Targeted trait	Marker system used	Markers found associated with trait	Phenotypic variance explained	References
Bulked segregants analysis (BSA) based							
GS1 × ICPL 87119	F_2	254	Fusarium wilt	RAPD	OPM03704, OPAC11500	–	Kotresh et al. (2006)
TT 44-4 × TDI 2004-1	F_2	84	Plant type	RAPD	$OPF04_{700}$, $OPA09_{1375}$	–	Dhanasekar et al. (2010)
TTB 7 × BRG 3	F_2	121	SMD resistance	AFLP	E-CAA/M-GTG_{150}, E-CAA/M-GTG_{60}	–	Ganapathy et al. (2009)
Genetic map and QTL based							
ICP 8863 × ICPL 20097	$F_{2:3}$	190	SMD resistance	SSR	CcM1982, CcM1447 (qSMD1)	9.2	Gnanesh et al. (2011)
					CcM0588, CcM2781 (qSMD2)	8.3	Gnanesh et al. (2011)
TTB 7 × ICP 7035	$F_{2:3}$	130	SMD resistance	SSR	CcM2149, CcM0468 (qSMD3)	12.32	Gnanesh et al. (2011)
					CcM1825, CcM1895 (qSMD4)	24.72	Gnanesh et al. (2011)
					CcM0970, CcM2485 (qSMD5)	15.93	Gnanesh et al. (2011)
					CcM0416, CcM2337 (qSMD6)	10.58	Gnanesh et al. (2011)
ICPA 2039 × ICPR 2447	F_2	188	Fertility restoration	SSR	CcM1522, CcM1821 (QTL-RF-1)	14.85	Bohra et al. (2012)
					CcM0047, CcM2332 (QTL-RF-2)	15.84	Bohra et al. (2012)
ICPA 2043 × ICPR 2671	F_2	188	Fertility restoration	SSR	CcM2542, CcM1277 (QTL-RF-3)	20.89	Bohra et al. (2012)
ICPA 2043 × ICPR 3467	F_2	188	Fertility restoration	SSR	CcM0374, CcM1506 (QTL-RF-4)	24.17	Bohra et al. (2012)
Pusa Dwarf × HDM04-1	$F_{2:3}$	186	Plant type and earliness	SSR and SNPs	ASNP1310-ASNP2099 (*q*PH4.1)	28.0	Kumawat et al. (2012)
					ASNP1310-ASNP2099 (*q*FL4.1)	51.4	Kumawat et al. (2012)
					ASNP1664-ASSR295 (*q*PB4.1)	19.5	Kumawat et al. (2012)
					ASSR100-ASSR206 (*q*SB5.1)	10.4	Kumawat et al. (2012)
					ASSR100-ASSR206 (*q*MT5.1)	25.9	Kumawat et al. (2012)
					ASSR100-ASSR206 (*q*PD5.1)	18.9	Kumawat et al. (2012)

In pigeonpea, extensive survey of BAC-end sequences (BESs) provided 3,072 BES-SSRs and all these BES-SSRs were further used for linkage analysis and trait mapping (Bohra et al. 2011, 2012; Gnanesh et al. 2011). In addition, a detailed microsatellite survey of whole genome sequence of pigeonpea has identified thousands of SSRs (Singh et al. 2011; Varshney et al. 2012).

In addition to SSRs, DArT offers great potential because of its sequence-independent nature and ensures whole genome profiling in a high throughput and cost effective manner. In pigeonpea, development of 5,376 DArT features helped in assessment of genetic diversity in a panel of 96 genotypes from 20 different *Cajanus* species (Yang et al. 2006). However, in the post genomics era, owing to the amenability to high throughput detection and precise genotyping, SNP has emerged as preferred class of DNA markers over SSRs. Thousands of SNPs were identified in pigeonpea to undertake genome wide association studies (GWAS) and genome wide selection (GWS) (Varshney et al. 2012; Saxena et al. 2012). Recently cost effective SNP genotyping assays such as competitive allele-specific polymerase chain reaction (KASPar) assays were developed for a total of 1,616 SNPs and designated as PKAMs (pigeonpea KASPar assay markers). Further utility of all these KASPar based SNPs were successfully demonstrated in genetic mapping and comparative analysis in pigeonpea (Saxena et al. 2012). In a similar instance 752 SNPs were successfully used to genotype a panel of 110 accessions (wild as well as cultivated) using GoldenGate assay and provided valuable evidences about gene flow, phylogeny and domestication bottlenecks occurred in pigeonpea (Kassa et al. 2012).

Furthermore, with an aim to leverage the DNA marker catalog, microarray-based markers such as single feature polymorphism (SFP) were also discovered for various parental combinations in pigeonpea. For example, the number of identified SFPs ranged from 780 to 854 between parents of several mapping populations. In total, a novel set of markers comprising 5,692 SFPs was reported (Saxena et al. 2011).

BAC Libraries

BAC libraries harbor large inserts of DNA ranging from 100 to 350 kb with an average insert size of 150 kb. The large size of DNA inserts ensures better coverage of the genome. These offer several advantages like ease of handling, high stability, non-chimeric nature and better transformation efficiency over other vectors such as yeast artificial chromosomes (YACs) and cosmids (Farrar and Donnison 2007). BAC libraries represent a potential genomic resource extensively used for (1) physical map construction, (2) comparative genome analysis via searching for macrosyntenic blocks across species, (3) map-based or positional cloning to isolate genes/QTLs responsible for economically important traits, (4) large scale DNA marker discovery through BAC-end sequencing, and (5) assembling of raw sequence reads into genome assembly for an organism. In pulses, several BAC libraries have been

reported and are being constructed for chickpea, lentil, pigeonpea, mungbean, cowpea, field pea and common bean etc. (Yu 2012). In pigeonpea, two BAC libraries were constructed by using *Hind*III and *Bam*HI restriction enzymes. Each of the libraries was composed of 34,560 clones. The average insert size of *Hind*III library was 120 kb while the *Bam*HI library had an average insert size of 115 kb. These clones collectively represented ~11× coverage of the pigeonpea genome. The sequences adjacent to the insertion sites are generally known as BESs and potential resources for identifying minimally overlapping clones (Kelley et al. 1999). With this perspective, randomly selected 50,000 BAC clones were targeted for end sequencing which generated a set of 88,860 high quality BESs (Bohra et al. 2011).

Genetic Maps

Saturated genetic maps have been constructed for several legumes like chickpea (Thudi et al. 2011; Hiremath et al. 2012), cowpea (Muchero et al. 2009; Lucas et al. 2011), common bean (Cordoba et al. 2010), soybean (Hwang et al. 2009) etc. Till 2010, no genetic map was available for pigeonpea due to non-availability of ample amount of genomic resources such as molecular markers and segregating mapping populations and this situation exacerbated by low genetic variation in *Cajanus* primary gene pool. Following the large scale development of BES-SSR and DArT markers, the first generation genetic maps were constructed for an F_2 population derived from an inter-specific cross ICP 28 (*C. cajan*) × ICPW 94 (*C. scarabaeoides*). SSR based genetic map covered a total map length of 930.9 cM with 239 loci with an average inter-marker distance of 3.8 cM (Bohra et al. 2011). In parallel, DArT based genotyping on this parental combination provided a set of 388 polymorphic markers. However, coupling and repulsion phase of polymorphic markers resulted in development of paternal and maternal specific genetic maps with 172 and 122 unique loci, respectively.

The above mentioned genetic maps were derived between *C. cajan* and *C. scarabaeoides*, which does not reflect the level of DNA polymorphism existing in primary or cultivated gene pool of *Cajanus*. Therefore, greater emphasis was directed towards construction of genetic maps for narrow crosses/intra-specific mapping populations. Keeping this view in mind, a total of six SSRs based intra-specific genetic maps with low to moderate marker density were constructed for six F_2 mapping populations (Gnanesh et al. 2011; Bohra et al. 2012). The number of mapped loci were in the range of 59 (ICPB 2049 × ICPL 99050) to 140 (ICPA 2043 × ICPR 3467) while covering map distances of 466.97 cM (TTB 7 × ICP 7035) to 881.57 cM (ICPA 2043 × ICPR 3467). Furthermore, to make the available linkage information more useful, all the six intra-specific genetic maps were joined together into a single consensus genetic map providing map positions to a total of 339 SSR markers at map coverage of 1,059 cM (Bohra et al. 2012). The bin wise polymorphism information content (PIC) values provided for each mapped loci will help geneticists and breeders to select the more informative and precise markers from the region of

interest. Recently one more genetic map based on an intra-specific mapping population (Pusa Dwarf × HDM04-1) was reported for cultivated pigeonpea. This genetic map comprising 296 loci (267 SNPs + 29 SSRs) covered a map length of 1520.22 cM organized into 11 LGs (Kumawat et al. 2012).

Inter-specific mapping population (ICP 28 × ICPW 94) relatively bigger than previously used (167 F_2) mapping populations, used for SNP genotyping through cost effective genotyping platform (KASPar assays) resulted in a much lower genotyping error rate than that obtained with markers like SSRs. A comprehensive genetic map comprising of 875 SNP loci was constructed (Saxena et al. 2012). The total length of this map was 967.03 cM with average marker distance of 1.11 cM. This linkage map was a considerable improvement with the previous pigeonpea genetic linkage maps using SSR and DArT markers. The marker density in this map has almost three times higher than the previous maps. This higher marker density would be useful in determining double recombinants affecting a single marker and in guiding future mapping efforts in pigeonpea.

Trait Mapping

Trait mapping is one of the important pre-requisite for prediction of phenotype from the genotype. As compared to some other legumes like chickpea and common bean not much progress has been witnessed in the area of trait mapping in pigeonpea. Earlier inadequate supply of DNA polymorphisms and lack of saturated genetic maps have posed obstacles in undertaking QTL analysis in pigeonpea. Despite this, some of the traits such as tolerance to SMD and FW and ideal plant type were chosen for mapping using bulked segregants analysis (BSA). BSA was performed using DNA from extremes phenotypes from segregating F_2 populations. The first instance of QTL analysis was reported by Gnanesh et al. (2011) to tag SMD resistance in pigeonpea. This study reported existence of major as well as minor effect QTLs imparting resistance against SMD. The investigation included two F_2 mapping populations which were subjected to linkage and QTL analysis. The results indicated occurrence of six QTLs (designated as qSMD1-6) explaining phenotypic variations in the range of 8.3–24.72 % (Gnanesh et al. 2011) (Table 5.2).

Another successful attempt for mapping a trait using QTL analysis was performed for fertility restoration (Rf). Restoration of fertility in hybrids forms a vital part of CMS based hybrid breeding. Keeping this in mind, QTL analysis was conducted using genotyping and phenotyping data generated from three different F_2 mapping populations showing segregation for fertility restoration. QTL analysis revealed a total of four large effect Rf-QTLs playing important roles in fertility restoration in pigeonpea (Table 5.2). The phenotypic variations governed by the identified QTLs were observed up to 24 % (Bohra et al. 2012). The SSR markers tightly linked with fertility restoration will help not only in search of a potential restorer but also in discriminating between restorer and maintainer. Similarly based on an intra-specific F_2 population and $F_{2:3}$ families (Pusa Dwarf × HDM04-1) several QTLs were

recovered for six different agronomics traits related to plant type and earliness and the phenotypic variation was observed in the range of 3–50 % (Kumawat et al. 2012). These genomic regions can further be chosen as candidates while practicing marker assisted selection (MAS) for plant type and earliness in pigeonpea.

Functional and Comparative Genomics

Functional genomics has shown remarkable impacts on plant genetics and breeding. In the context, collection of ESTs represents an excellent genomic resource to carry out functional genomics studies. In pigeonpea, a total of 25,576 ESTs have been deposited in NCBI database (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). In parallel, recent advancements made in the area of next generation sequencing technologies have helped generation of massive transcriptome sequence data. For instance, in the case of pigeonpea, a total of four transcriptome assemblies have been constructed using Illumina GA IIx, FLX/454 and Sanger sequencing (Raju et al. 2010; Dutta et al. 2011; Dubey et al. 2011; Kudapa et al. 2012). Among these, the two most comprehensive assemblies were designated as *Cajanus cajan* transcriptome assembly version 1 and 2 (CcTA v1: Dubey et al. 2011 and CcTA v2: Kudapa et al. 2012) comprising 48,726 and 21,434 transcript assembly contigs (TACs), respectively. The robust transcriptome assembly offers tremendous scope for predicting gene content, function and large scale mining of genic or functional molecular markers (GMM or FMM). For instance, different sets of EST-SSRs were developed from these transcriptome assemblies and validated in a panel of diverse pigeonpea genotypes (Raju et al. 2010; Dutta et al. 2011). Since the functional markers remain highly conserved across genera during the course of evolution, these form the basis for comparative genome analysis.

Comparative genomics remains a powerful approach to harness genomic information form related genera. In pigeonpea, successful cloning of approximately 600 unique nucleotide-binding site (NBS) domain and leucine-rich repeat (LRR) domain sequences was performed using degenerate primers targeting the NBS-LRR sequences from model legume *Medicago truncatula* (Varshney et al. 2010a). NBS-LRR represents the most abundant class of resistance genes in plants (Varshney et al. 2009). Therefore, availability of cloned NBS-LRR fragments would shed light into the fate of NBS-LRR resistance genes during divergence of Millettioid and Galegoid clades within the subfamily *Papilionoideae*. Similarly, comparative analysis of the CcTA v2 with genome sequence of soybean (*Glycine max*) provided a set of 128 intron spanning region (ISR) markers. Mapped SNPs were also used to discover the synteny blocks in each of the 11 pigeonpea linkage groups to their counterparts of four legumes chromosomes (soybean, cowpea, *Medicago* and *Lotus*), implying certain co-linearity for the syntenic chromosome/linkage pairs. Conserved sequences were identified among five legume species (pigeonpea, soybean, cowpea, *Medicago* and *Lotus*) (Saxena et al. 2012). The data from comparative genome analysis should facilitate studies on genome evolution and analysis of structural genome, but more significantly would be helpful in understanding and validation

of functional inference of genes in pigeonpea. The identification of gene functions is difficult in non-model species including pigeonpea, thus functional genome analysis will have to rely heavily on the establishment of orthologies from model species by using comparative genomics analysis.

Genome Sequencing

With the availability of draft genome sequence, pigeonpea has shown a quantum jump in its status and joined the league of model/genomic resource rich crops. Pigeonpea has become the first orphan and non-industrial grain legume to have a draft genome sequence (Varshney et al. 2012). Next generation sequencing platforms such as Illumina GA and HiSeq 2000 were used to sequence elite pigeonpea cultivar Asha (ICPL 87119). Using a *de novo* genome assembly and with the help of bacterial artificial chromosome (BAC)-end sequences and available genetic maps, 605.78 Mb was assembled into scaffold with N50 = 516.06 kb. Based on estimated genome size of 833 Mb using a K-mer analysis, 72.8 % of the genome was assembled. Gene prediction analysis suggested presence of 48,680 genes with an average transcript length of 2,348 bp and 3.59 exons per gene. A total of 46,750 genes (96.4 %) could be assigned based on functional ontology and 1,930 genes (3.96 %) are of unknown function. In terms of non-coding RNAs (ncRNAs), 763 tRNA, 862 miRNA, 329 rRNA and 363 snRNA were encountered in <0.5 % assembled genome. In another sequencing effort, 454 GS-FLX sequencing technology was used to assemble ~511 Mb sequence data for Asha variety (Singh et al. 2011). In this study, 47,004 protein coding genes including 1,213 disease resistance/defence related genes and 152 abiotic stress tolerance genes were predicted.

Analysis of genome assembly (Varshney et al. 2012) for repetitive DNA showed presence of transposable elements (TEs) in 49.61 % of assembled genome. Comparison of pigeonpea genome with soybean, grape, *Medicago truncatula* and *Lotus japonicus* genomes revealed 4,311 clusters of genes that were common to all five eudicot genomes whereas 3,068 gene families were specific to the pigeonpea genome. Pigeonpea genome contains higher number (111) of drought responsive genes than soybean, *Medicago truncatula* and *Lotus japonicus*. These genes are suitable candidates for allele mining in global germplasm collection of pigeonpea so that superior alleles and haplotypes for drought tolerance can be implemented in pigeonpea crop improvement (Varshney et al. 2012).

Genome sequence will be useful in utilizing gene sequences for molecular breeding as well as genetic engineering approaches for crop improvement to minimize yield gap in farmers' field. It will not only facilitate comparative analysis with other members of warm-season Millettioids and cool-season Galegoids but also in understanding the phylogeny and evolution within the legume family as a whole. Furthermore, identified candidate drought responsive genes can be utilized for improving other legume crops such as soybean and common bean, which are adversely affected by drought stress.

Genomics-Assisted Breeding

To enhance the crop productivity of pigeonpea, it is important to implement recently developed biotechnological tools such as molecular markers and genetic maps in the breeding programs. These are pre-requisites for genomics-assisted breeding applications such as marker-assisted selection (MAS) (Varshney et al. 2009). With the development and availability of molecular markers and dense molecular genetic maps, MAS is in routine in breeding programs in several major crop species. However, in pigeonpea full potential of molecular breeding still needs to be realized to reap the benefits of colossal amount of molecular information generated through whole genome sequencing. Though, traditional pigeonpea breeding has provided enough genetic gains in the form of release of several elite cultivars, the pace of improvement is not adequate. Wild relatives of pigeonpea representing the untapped reservoir of tremendous genetic variation offer greater scope for broadening of genetic base in pigeonpea. However the undesirable alleles associated with the wild germplasm i.e. linkage drag hampers the speedy recovery of superior performance. Some novel molecular breeding methods such as advanced backcross QTL (AB-QTL) analysis permitting identification as well as transfer of wild type superior alleles into elite cultivars help greatly by generating broad based breeding materials including introgression lines (ILs), near isogenic lines (NILs), chromosome segment substitution lines (CSSLs) etc. Some efforts have also been initiated at ICRISAT using *C. scarabaeoides* as donor to discover superior alleles of various economically important QTLs through AB-QTL approach (Varshney et al. 2013).

Apart from this, whole genome opens new avenues for re-sequencing and genome wide SNP genotyping of landraces/core/reference sets/composite collection (Upadhyaya et al. 2011) (Fig. 5.1). This will greatly assist in discovery of alleles and unlocking the alleles/loci undergoing selection pressure during the process of domestication. In addition, reference genome would facilitate precise identification of recombination blocks using high throughput genotyping platforms and methods such as genotyping by sequencing (GBS). GBS can be employed to tap the potential of nested association mapping (NAM) ensuring benefits of both association mapping (historical recombination) as well as linkage analysis (bi-parental recombination). NAM would provide insights into the molecular basis underlying various QTLs governing several complex traits. In crops like pigeonpea, some of the other schemes relying on multi-parent crossing would be very effective in providing opportunities for extensive recombination. For instance, creation of multi-parent advanced generation intercross (MAGIC) lines in pigeonpea will help not only in accumulation of superior alleles from various genetic backgrounds but also in fine mapping of the region of interest (Kover et al. 2009). Access to the genome wide SNP/SSR markers together with availability of a training population with a robust historical phenotyping data would allow identification of a genotype with higher breeding value through genomic selection (GS) bypassing extensive field testing/repeated phenotyping (Varshney et al. 2013).

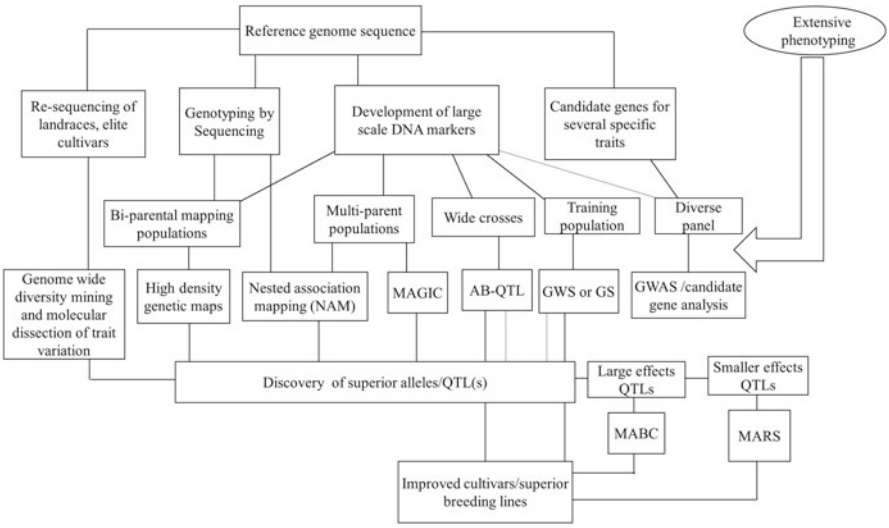


Fig. 5.1 An integrated approach to harness reference genome sequence for pigeonpea genetic information

Conclusion and Perspectives

With the availability of draft genome sequence, pigeonpea has marked its presence among sequenced legumes such as *Lotus*, *Medicago* and soybean enabling more focus on basic research and translational genomics for crop improvement. In the context, high density genetic maps along with precise phenotyping platforms would facilitate identification of genomic regions/QTLs associated with traits such as tolerance to abiotic and biotic stress and fertility restoration. Since exploitation of hybrid vigor seems to a potential alternative to overcome the existing yield barriers, elaborated understanding about the molecular basis of heterosis would allow easy access to the genes imparting hybrid vigor. Furthermore, re-sequencing of several genotypes including landraces, wild relatives and cultivars would ensure recovery of novel haplotypes associated with domestication and other important phenomenon. The deployment of these genomic tools into regular breeding programmes in the form of MABC, MARS and GS would help greatly in bridging the yield gap in pigeonpea through enhancement of productivity in the resource poor and risk prone environment.

Acknowledgements Authors from ICRISAT are thankful to United States Agency for International Development (USAID) for financial support to utilize genome sequence information for pigeonpea improvement. This work has been undertaken as part of the CGIAR Research Program on Grain Legumes.

References

Akinola JO, Pritchard AJ, Whiteman PC (1972) Chromosome number in pigeonpea (*Cajanus cajan* (L.) Millsp.). J Aust Inst Agric Sci 38:305

Bohra A, Dubey A, Saxena RK, Penmetsa RV, Poornima KN, Kumar N, Farmer AD, Srivani G, Upadhyaya HD, Gothalwal R, Ramesh R, Singh D, Saxena KB, KaviKishor PB, Singh NK, Town CD, May GD, Cook DR, Varshney RK (2011) Analysis of BAC-end sequences (BESs) and development of BES-SSR markers for genetic mapping and hybrid purity assessment in pigeonpea. BMC Plant Biol 11:56

Bohra A, Saxena RK, Gnanesh BN, Saxena KB, Byregowda M, Rathore A, KaviKishor PB, Cook DR, Varshney RK (2012) An intra-specific consensus genetic map of pigeonpea [*Cajanus cajan* (L.) Millspaugh] derived from six mapping populations. Theor Appl Genet 125:1325–1338

Burns MJ, Edwards KJ, Newbury HJ, Ford-Lloyd BV, Baggott CD (2001) Development of simple sequence repeat (SSR) markers for the assessment of gene flow and genetic diversity in pigeonpea (*Cajanus cajan*). Mol Ecol Notes 1:283–285

Cordoba JM, Chavarro C, Schlueter JA, Jackson SA, Blair MW (2010) Integration of physical and genetic maps of common bean through BAC-derived microsatellite markers. BMC Genomics 11:436

Dhanasekar P, Dhumal KH, Reddy KS (2010) Identification of RAPD marker linked to plant type gene in pigeonpea. Indian J Biotechnol 9:58–63

Dubey A, Farmer A, Schlueter J, Cannon SB, Abernathy B, Tuteja R, Woodward J, Shah T, Mulasmanovic B, Kudapa H, Raju NL, Gothalwal R, Pande S, Xiao Y, Town CD, Singh NK, May GD, Jackson S, Varshney RK (2011) Defining the transcriptome assembly and its use for genome dynamics and transcriptome profiling studies in pigeonpea (*Cajanus cajan* L.). DNA Res 18:153–164

Dutta S, Kumawat G, Singh BP, Gupta DK, Singh S, Dogra V, Gaikwad K, Sharma TR, Raje RS, Bandhopadhya TK, Datta S, Singh MN, Bashasab F, Kulwal P, Wanjari KB, Varshney RK, Cook DR, Singh NK (2011) Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. BMC Plant Biol 11:17

FAOSTAT.database(2011)http://faostat.fao.org/site/567/DextopDEfault.aspx?PageID=567#ancor. Accessed 14 Mar 2013

Farrar K, Donnison IS (2007) Construction and screening of BAC libraries made from *Brachypodium* genomic DNA. Nat Protoc 2:1661–1674

Ganapathy KN, Byregowda M, Venkatesha SC, Rama Chandra R, Gnanesh BN, Girish G (2009) Identification of AFLP markers linked to sterility mosaic disease in pigeonpea *Cajanus cajan*(L.) Millsp. Int J Integr Biol 7:145–149

Gnanesh BN, Bohra A, Sharma M, Byregowda M, Pande S, Wesley V, Saxena RK, Saxena KB, KaviKishor PB, Varshney RK (2011) Genetic mapping and quantitative trait locus analysis of resistance to sterility mosaic disease in pigeonpea [*Cajanus cajan* (L.) Millsp.]. Field Crops Res 123:53–61

Hiremath PJ, Kumar A, Penmetsa RV, Farmer A, Schlueter JA, Chamarthi SK, Whaley AM, Carrasquilla-Garcia N, Gaur PM, Upadhyaya HD, KaviKishor PB, Shah TM, Cook DR, Varshney RK (2012) Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. Plant Biotechnol J 10:716–732

Hwang TY, Sayama T, Takahashi M, Takada Y, Nakamoto Y, Funatsuki H, Hisano H, Sasamoto S, Sato S, Tabata S, Kono I, Hoshi M, Hanawa M, Yano C, Xia Z, Harada K, Kitamura K, Ishimoto M (2009) High-density integrated linkage map based on SSR markers in soybean. DNA Res 16:213–225

Kassa MT, Penmetsa RV, Carrasquilla-Garcia N, Sarma BK, Datta S, Upadhyaya HD, Varshney RK, Von Wettberg EJB, Cook DR (2012) Genetic patterns of domestication in pigeonpea (*Cajanus cajan* (L.) Millsp.) and wild *Cajanus* relatives. PLoS One 7:e39563

Kelley JM, Field CE, Craven MB, Bocskai D, Kim UJ, Rounsley SD, Adams MD (1999) High throughput direct end sequencing of BAC clones. Nucleic Acids Res 27:1539–1546

Kotresh H, Fakrudin B, Punnuri S, Rajkumar B, Thudi M, Paramesh H, Lohithswa H, Kuruvinashetti M (2006) Identification of two RAPD markers genetically linked to a recessive allele of a *Fusarium* wilt resistance gene in pigeonpea (*Cajanus cajan* (L.) Millsp.). Euphytica 149:113–120

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM et al (2009) A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. PLoS Genet 5:e1000551

Krishnaswamy N, Ayyangar GNR (1935) Chromosome number in *Cajanus indicus* Spreng. Curr Sci 3:614

Kudapa H, Bharti AK, Cannon SB, Farmer AD, Mulaosmanovic B, Kramer R, Bohra A, Weeks NT, Crow JA, Tuteja R, Shah T, Dutta S, Gupta DK, Singh A, Gaikwad K, Sharma TR, May GD, Singh NK, Varshney RK (2012) A comprehensive transcriptome assembly of pigeonpea (*Cajanaus cajan* L.) using Sanger and Second-generation sequencing platforms. Mol Plant 5:1020–1028

Kumawat G, Raje RS, Bhutani S, Pal JK, Mithra SVCR, Kishor Gaikwad K, Sharma TR, Singh NK (2012) Molecular mapping of QTLs for plant type and earliness traits in pigeonpea (*Cajanus cajan* L. Millsp.). BMC Genet 13:84

Lucas MR, Diop NN, Wanamaker S, Ehlers JD, Roberts PA, Close TJ (2011) Cowpea soybean synteny clarified through an improved genetic map. Plant Genome 4:218–225

Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S, Pottorff M, Hearne S, Cisse N, Fatokun C, Ehlers JD, Roberts PA, Close TJ (2009) A consensus genetic map of cowpea [*Vigna unguiculata* (L) Walp.] and synteny based on EST-derived SNPs. Proc Natl Acad Sci U S A 106:18159–18164

Naithani SP (1941) Cytological studies on Indian pulses, Part 1, the somatic chromosomes and the pro-chromosomes of *Cajanus*. Proc Natl Acad Sci India 11:67

Odeny DA, Jayashree B, Ferguson M, Hoisington D, Crouch J, Gebhardt C (2007) Development, characterization and utilization of microsatellite markers in pigeonpea. Plant Breed 126:130–136

Odeny DA, Jayashree B, Gebhardt C, Crouch J (2009) New microsatellite markers for pigeonpea (*Cajanus cajan* (L.) Millsp.). BMC Res Notes 2:35

Panguluri SK, Janaiah J, Govil JN, Kumar PA, Sharma PC (2005) AFLP fingerprinting in pigeonpea (*Cajanus cajan* L. Millsp.) and its wild relatives. Genet Res Crop Evol 53:523–531

Raju NL, Gnanesh BN, Lekha P, Jayashree B, Pande S, Hiremath PJ, Byregowda M, Singh NK, Varshney RK (2010) The first set of EST resource for gene discovery and marker development in pigeonpea (*Cajanus cajan*L.). BMC Plant Biol 10:45

Ratnaparkhe MB, Gupta VS, Ven Murthy MR, Ranjekar PK (1995) Genetic fingerprinting of pigeonpea (*Cajanus cajan* (L.) Millsp) and its wild relatives using RAPD markers. Theor Appl Genet 91:893–898

Saxena RK, Prathima C, Saxena KB, Hoisington DA, Singh NK, Varshney RK (2010a) Novel SSR markers for polymorphism detection in pigeonpea (*Cajanus* spp.). Plant Breed 129:142–148

Saxena RK, Saxena KB, Varshney RK (2010b) Application of SSR markers for molecular characterization of hybrid parents and purity assessment of ICPH 2438 hybrid of pigeonpea [*Cajanus cajan* (L.) Millspaugh]. Mol Breed 26:371–380

Saxena RK, Cui X, Thakur V, Walter B, Close TJ, Varshney RK (2011) Single feature polymorphisms (SFPs) for drought tolerance in pigeonpea (*Cajanus* spp.). Funct Integr Genomics 11:651–657

Saxena RK, Penmetsa RV, Upadhyaya HD, Kumar A, Carrasquilla-Garcia N, Schlueter JA, Farmer A, Whaley AM, Sarma BK, May GD, Cook DR, Varshney RK (2012) Large-scale development of cost-effective single-nucleotide polymorphism marker assays for genetic mapping in pigeonpea and comparative mapping in legumes. DNA Res 19:449–461

Singh NK, Gupta DK, Jayaswal PK, Mahato AK, Dutta S, Singh S, Bhutani S, Dogra V, Singh BP, Kumawat G, Pal JK, Pandit A, Singh A, Rawal H, Kumar A, Prashat RG, Khare A, Yadav R,

Raje RS, Singh MN, Datta S, Fakrudin B, Wanjari KB, Kansal R, Dash PK, Jain PK, Bhattacharya R, Gaikwad K, Mohapatra T, Srinivasan R, Sharma TR (2011) The first draft of the pigeonpea genome sequence. J Plant Biochem Biotechnol 21:98–112

Sivaramakrishnan S, Seetha K, Rao AN, Singh L (1997) RFLP analysis of cytoplasmic male sterile lines in pigeonpea (*Cajanus cajan* L. Millsp.). Euphytica 126:293–299

Sivaramakrishnan S, Seetha K, Reddy LJ (2002) Diversity in selected wild and cultivated species of pigeonpea using RFLP of mtDNA. Euphytica 125:21–28

Thudi M, Bohra A, Nayak SN, Varghese N, Shah TM, Penmesta RV, Thirunavukkarasu N, Gudipati S, Gaur PM, Kulwal PL, Upadhyaya HD, Kavikishor PB, Winter P, Kahl G, Town CD, Kilian A, Cook DR, Varshney RK (2011) Novel SSR markers from BAC-end sequences, DArT arrays and a comprehensive genetic map with 1,291 marker loci for chickpea (*Cicer arietinum* L.). PLoS One 6:e27275

Upadhyaya HD, Reddy KN, Sharma S, Varshney RK, Bhattacharjee R, Singh S, Gowda CLL (2011) Pigeonpea composite collection and identification of germplasm for use in crop improvement programmes. Plant Genet Resour 9:97–108

Varshney RK, Close TJ, Singh NK, Hoisington DA, Cook DR (2009) Orphan legume crops enter the genomics era. Curr Opin Plant Biol 12:1–9

Varshney RK, Penmetsa RV, Dutta S, Kulwal PL, Saxena RK, Datta S, Sharma TR, Rosen B, Carrasquilla-Garcia N, Farmer AD, Dubey A, Saxena KB, Gao J, Fakrudin B, Singh MN, Singh BP, Wanjari KB, Yuan M, Srivastava RK, Kilian A, Upadhyaya HD, Mallikarjuna N, Town CD, Bruening GE, He G, May GD, McCombie R, Jackson SA, Singh NK, Cook DR (2010a) Pigeonpea genomics initiative (PGI): an international effort to improve crop productivity of pigeonpea (*Cajanus cajan* L.). Mol Breed 26:393–408

Varshney RK, Glaszmann JC, Leung H, Ribaut JM (2010b) More genomic resources for less-studied crops. Trends Biotechnol 28:452–460

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MTA, Azam S, Fan G, Whaley AM, Farmer AD, Sheridan J, Iwata A, Tuteja R, Penmetsa RV, Wu W, Upadhyaya HD, Yang SP, Shah T, Saxena KB, Michael T, McCombie WR, Yang B, Zhang G, Yang H, Wang J, Spillane C, Cook DR, May GD, Xu X, Jackson SA (2012) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotechnol 30:83–89

Varshney RK, Murali Mohan S, Gaur PM, Gangarao NVPR, Pandey MK, Bohra A, Sawargaonkar SL, Kimurto PK, Janila P, Saxena KB, Fikre A, Sharma M, Pratap A, Tripathi S, Datta S, Chaturvedi SK, Mallikarjuna N, Anuradha G, Babbar A, Choudhary AK, Mhase MB, Bharadwaj CH, Mannur DM, Harer PN, Guo B, Liang X, Nadarajan N and Gowda CLL (2013) Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. Biotechnol Adv S0734-9750(13)00003-7

Yang S, Pang W, Harper J, Carling J, Wenzl P, Huttner E, Zong X, Kilian A (2006) Low level of genetic diversity in cultivated pigeonpea compared to its wild relatives is revealed by diversity arrays technology (DArT). Theor Appl Genet 113:585–595

Yang SS, Xu WW, Tesfaye M, Lamb JAFS, Jung HJG, Samac DA, Vance CP, Gronwald JW (2009) Single-feature polymorphism discovery in the transcriptome of tetraploid alfalfa. Plant Genome 2:224–232

Yang SY, Saxena RK, Kulwal PL, Ash GJ, Dubey A, Harpe JDI, Upadhyaya HD, Gothalwal R, Kilian A, Varshney RK (2011) The first genetic map of pigeonpea based on diversity arrays technology (DArT) markers. J Genet 90:103–109

Yu K (2012) Bacterial chromosome libraries of pulse crops—characteristics and applications. J Biomed Biotechnol 2012:493186

Chapter 6
Advances in Lentil Genomics

Shiv Kumar, Aladdin Hamwieh, Alagu Manickavelu, Jitendra Kumar, Tilak Raj Sharma, and Michael Baum

Abstract Lentil is a diploid (2n = 2X = 14) self-pollinating crop with a genome size of 4 Gbp. The use of genomics tools in lentil breeding programs has been limited, since available genomic resources are not adequate. Recent advances in high-throughput genotyping and sequencing technologies have brought in new impetus in the development of genetic and genomic resources and high resolution marker-trait association in lentil. Their integration in marker-assisted breeding is expected to improve the precision and efficiency in breeding programs with accelerated and directed genetic gains in crops like lentil. Molecular markers are expected to facilitate indirect selection for difficult traits, introgression of novel genes into adapted varieties, pyramiding genes from different sources, and combining multiple stress resistance. The present review highlights recent advances in lentil genomics and future outlook in the light of rapid advancement in the genomics tools.

Keywords Lentil • *Lens culinaris* • Molecular marker • Genomic resources • Marker assisted breeding

S. Kumar, Ph.D. (✉)
Biodiversity and Integrated Gene Management (BIGM) Program,
International Center for Agricultural Research in the Dry Areas,
Rabat-Institut, ICARDA/Rabat Office, PO Box 6299, Rabat, Morocco
e-mail: sk.agrawal@cgiar.org

A. Hamwieh, Ph.D. • M. Baum, Ph.D.
Biodiversity and Integrated Gene Management (BIGM) Program, International Center for Agricultural Research in the Dry Areas (ICARDA), Aleppo, Syria

A. Manickavelu, Ph.D.
Biodiversity and Integrated Gene Management (BIGM) Program, International Center for Agricultural Research in the Dry Areas, Rabat, Morocco

J. Kumar, Ph.D.
Crop Improvement Division, Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

T.R. Sharma, Ph.D.
Department of Agricultural Biotechnology, CSK Himachal Pradesh Agricultural University, Palampur, Himachal Pradesh, India

S. Gupta et al. (eds.), *Legumes in the Omic Era*, DOI 10.1007/978-1-4614-8370-0_6,

Introduction

Lentil (*Lens culinaris* ssp. *culinaris* Medicus) is an important grain legume species cultivated throughout the West Asia, North Africa, the Indian subcontinent, North America, and Australia, providing a vital source of dietary protein in human diets and protein-rich straw for animal feed (Erskine et al. 2011). Lentil shares the ability to fix atmospheric nitrogen with other legumes, making it a useful option for soil fertility management in cereal based cropping systems. Lentil also provides rotational benefits to cereal crops in management of weeds, diseases and insect pests, and in many cases offers a profitable, high value crop option for farmers (Rahman et al. 2009). Genetic enhancement programs have been undertaken using conventional breeding approaches to improve yield and adaptability of the crop at national and global levels. Recently, deployment of genomic resources has become an integral part of breeding programs in many crops worldwide. However in comparison to cereal and major food legume crops, there are limited efforts in the development and deployment of molecular tools in lentil. Molecular tools have recently been used by lentil breeders and biotechnologists to understand the genetic basis of a few traits related to biotic (ascochyta blight, anthracnose, rust, fusarium wilt, stemphylium blight) and abiotic (drought, frost, cold, boron, salinity) stresses (Muehlbauer et al. 2006). However, in the genomic era, there is a need to keep pace with the development of new molecular tools and techniques such as transcriptomics and whole genome sequencing. Whole genome sequencing projects have been undertaken for model legumes like *Medicago* and *Lotus*, providing an opportunity to identify putative orthologous gene sequence resources in other legume species, especially those located within the Galegoid clade of the Fabaceae sub-family Papilionoideae. In addition, a draft genome sequence has recently been completed for the warm-season food legumes, soybean (Schmutz et al. 2010) and pigeon pea (Varshney et al. 2011), which belong to the Phaseoloid clade providing further insights into comparative genomics within the Fabaceae family.

In the present chapter, we reviewed lentil genomics, focusing on the present status of genomic resources, molecular markers, genetic engineering and the future outlook in the light of rapid advancement in the genomics tools.

Genome Size

Lentil is a diploid (2n = 2X = 14) self-pollinating crop with a genome size of 4 Gbp (Arumuganathan and Earle 1991). Variable chromosome numbers have been reported in inter-specific hybrids e.g., seven bivalents in the intra-specific hybrids within the members of *L. culinaris* to five bivalents and one quadrivalent in the F_1 hybrids derived from a cross between *L. culinaris* and *L. orientalis*. Several workers have studied karyotypes in *L. culinaris*, and reported similar karyotypes. The length of chromosomes ranged from 3.0 to 9.2 μm. Gupta and Singh (1981) reported two

pairs each of metacentric and sub-metacentric and three pairs of acrocentric chromosomes. The large size of nuclear genome in lentil poses a great challenge to sequence it with the limited resources.

Genomic Resources

Genomic resources are very important for a crop improvement program. Molecular approaches have made limited progress in improving the understanding of the lentil genome. Lentil breeding programs do not yet use marker-assisted selection (MAS) because the genetic maps developed in lentil with restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD) and simple sequence repeat (SSR) markers are not tightly linked to the genes of interest. Therefore, their use in identification of close trait-marker association remains a major limitation in lentil. In the recent past, however, efforts have been augmented towards the development of genomic resources in lentil.

Microsatellite or SSR Markers

Recently, major efforts have been directed towards the development of microsatellite and gene-specific markers in lentil. Microsatellite or SSR markers are generally co-dominant, unilocus, multi-allelic and species specific. Produced from primers designed to the flanking sequence of mainly di- and trinucleotide repeats, microsatellites have emerged the markers of choice in many plant species to gain understanding of genetic relationships, evolutionary insights and gene mapping. Development of microsatellite markers requires a considerable amount of laboratory effort. Successful isolation of microsatellite markers involves construction and screening of small insert genomic library with SSR motifs, sequencing of the positive clones, designing the primers that can amplify SSR loci, and determining polymorphic SSR primers. The first *Sau*3AI genomic library was constructed from the cultivated accession, ILL5588 and screened with (GT)10, (GA)10, (GC)10, (GAA)8, (TA)10 and (TAA) probes (Hamwieh et al. 2005). Dinucleotide repeats were observed more frequently than trinucleotide repeats or other motifs (Table 6.1). The microsatellite motifs were classified as perfect, imperfect, compound perfect or compound imperfect repeats according to the modified classification of Weber (1990). The simple/perfect repeats were predominant (56.8 %) followed by compound/perfect (16.1 %) whereas compound/imperfect (12.7 %) occurred least often. Among the perfect repeats, (CA/GT)n motifs were the most abundant, comprising 24.2 % of the isolated clones, followed by (AT/TA)n repeats (8.9 %). This led to a set of 30 highly polymorphic SSR markers in lentil. Hamwieh et al. (2009) further developed an additional set of 14 microsatellite markers and used them for genetic diversity analysis of the lentil

Table 6.1 Microsatellite motifs observed in the lentil genomic library

Type		Microsatellite motif	Number	% Occurrence
Simple	Perfect	CA/GT	57	24.2
		CG/GC	2	0.8
		CT/GA	7	3
		CTT/GAA	3	1.3
		AT/TA	21	8.9
		ATT/TAA	7	3
		Others types	37	15.7
		Total	134	56.8
	Imperfect	CA/GT	21	8.9
		CG/GC	0	0
		CT/GA	1	0.4
		CTT/GAA	0	0
		AT/TA	3	1.3
		ATT/TAA	1	0.4
		Others types	8	3.4
		Total	34	14.4
Compound	Perfect		38	16.1
	Imperfect		30	12.7
	Total		68	28.8
Total			236	100

Based on data from Hamwieh A, Udupa SM, Sarker A, Jung C and Baum M. Development of new microsatellite markers and their application in the analysis of genetic diversity in lentils. Breed Sci 2009; 59:77–86

core set. Recently, 126 SSR markers were generated using a magnetic bead capture method at the Washington State University (Weidong Chen and P.N. Rajesh, Personal communication). More than 500 SSRs have been generated from enriched genomic library in lentil (Sabhyata Bhatia, Personal Communication). Still, the lentil-specific SSR markers are limited and not sufficient for the genome-wide coverage to establish genetic relatedness among the closely related germplasm accessions.

Expressed Sequence Tags (ESTs)

ESTs are short DNA sequences of 150–400 bp from a cDNA clone that corresponds to a particular mRNA. These are developed and publicly made available (Rudd 2003). Development of high-throughput functional genomics approaches like Serial Analysis of Gene Expression (SAGE) has led to the generation of more ESTs. The cDNA clones corresponding to the ESTs of interest can be used as RFLP or CAPS based markers (Varshney et al. 2005). The EST sequence data also serve the purpose of identifying SSRs and/or SNPs. Before the ESTs, development of SSR and

SNP markers was expensive and required high resource laboratories, but presently any user can download them from the database and use some special bioinformatic programs like MISA for SSR detection (Thiel et al. 2003; Varshney et al. 2005) and SNipPer for SNP discovery (Kota et al. 2003; Varshney et al. 2005). As on October 2012 there are about 10,163 ESTs available for lentil. Most of the available ESTs (9,513) were published in November 2010 and another remaining 647 ESTs in September 2012 by the University of Saskatchewan. The first EST library was made from a mixture of eight cultivars with varying seed phenotypes. The second cDNA library was prepared from the leaflets of a Canadian cultivar 'Eston' inoculated with *Colletotrichum truncatum*. Kaur et al. (2011) carried out transcriptome sequencing of lentil based on the second-generation technology which permits large-scale unigene assembly and SSR marker discovery. They used tissue-specific cDNA samples from six genotypes (Northfield, ILL2024, Indianhead, Digger, ILL6788, and ILL7537) using Roche 454 GS-FLX Titanium technology, and generated c. 1.38×106 ESTs. *De novo* assembly generated 15,354 contigs and 68,715 singletons. Out of huge ESTs produced, 3,470 SNP and EST-SSRs have been identified. Development of genomic resources has become possible and cost effective with the advent of next generation sequencing of ESTs. Validation of a subset of 192 EST-SSR markers across a panel of 12 cultivated genotypes showed 47.5 % polymorphism from a set of 2,393 EST-SSR markers developed in lentil (Kaur et al. 2011).

Single Nucleotide Polymorphism (SNP)

SNP markers are considered ideal for genetic mapping and diversity assessment in crop plants due to their abundance and relatively even distribution across the genome (Chagne et al. 2007). In addition various technologies exist for evaluation of SNP loci and many of these are amenable to automation for allele calling and data collection. In fact, the availability of extensive sequence database made a new beginning to exploit them as a high-throughput marker system for genome mapping studies. The availability of abundant high-throughput sequence-based markers is essential for detailed genome-wide trait analysis. A significant amount of efforts has been invested in re-sequencing alleles to discover SNPs. There are techniques to detect SNPs such as allele-specific PCR, single base extension and array hybridization methods. Since SNP discovery and genotyping require expensive and sophisticated platforms, the development and exploitation of SNP markers are still restricted to major crop species such as rice (Nasu et al. 2002), wheat (Somers et al. 2003), barley (Kota et al. 2001; Kanazin et al. 2002), maize (Tenaillon et al. 2001), and soybean (Zhu et al. 2003). Recent advances in sequencing techniques have helped in developing SNP assay in lentil. The Department of Primary Industries (Victoria, Australia) and University of Saskatchewan are involved in developing lentil genomic resource by developing SNP and COS marker assay.

Marker Transferability

Comparative genome mapping has demonstrated different levels of genome conservation among crop species during the course of evolution (Choi et al. 2004; Zhu et al. 2005). The lentil genome has shown different degrees of synteny with other legume crops (Weeden et al. 1992; Simon and Muehlbauer 1997; Phan et al. 2007; Choudhary et al. 2008). Development of PCR-based markers has improved transferability of genetic information among species through comparative genomics, and has facilitated the establishment of phylogenetic relationship in plants species. Since the availability of microsatellite markers in lentil is limited, other legumes offer great scope of marker transferability for genome-wide coverage. Pandian et al. (2000) observed 5 % transferability of chickpea-specific STMS primers in lentil while Reddy et al. (2009) observed successful amplification of 62 % *Trifolium* markers followed by *Medicago* (36 %) and *Pisum* (25 %). Datta et al. (2011) reported transferability of 19 STMS markers from common bean, chickpea, pigeon pea, and soybean. The lack of lentil-specific microsatellite sequences and gene-based markers propelled the mining and transfer of expressed sequence tag-simple sequence repeat (EST-SSR) sequences from the model genome *Medicago truncatula* to enrich an existing intraspecific lentil genetic map (Gupta et al. 2012a). They published 21 clear and reproducible markers showing polymorphism between parents, Northfield and Digger. EST-based intron-targeted amplified polymorphic (ITAP) markers have recently been developed from related crops and applied to lentil. ESTs were compared for phylogenetic distant from *M. truncatula*, *Lupinus albus*, and *G. max* to produce 500 ITAP markers that could be applied to lentil (Phan et al. 2007). Also, 126 *M. truncatula* cross-species markers were used to generate comparative genetic maps of lentil and white lupin and macrosyntenic relationships between lentil and field pea was observed.

Functional Genomics

Differential gene transcript profiles were assessed among resistant (ILL7537) and susceptible (ILL6002) lentil genotypes at 6, 24, 48, 72 and 96 h after inoculation (hai) with *Ascochyta lentis* (AL4 isolate) (Ford et al. 2007). The non-redundant differentially expressed genes for each accession and time point were hierarchically clustered using Euclidean metrics. In total, 25 differentially expressed sequences were up-regulated and 56 down-regulated in ILL7537 whereas 26 were up-regulated and 44 down-regulated in ILL6002. Several candidate defense genes were characterized from lentil including a *b*-1,3-glucanase, a pathogenesis-related protein from the Bet v I family, a pea disease resistance response protein 230 (DRR230-a), a disease resistance response protein (DRRG49-C), a PR4 type gene and a gene encoding an antimicrobial SNAKIN2 protein, all of which have been fully sequenced. Several transcription factors were also recovered at 6 hai and future

aims will be to further biologically characterize these and earlier responses to gain a comprehensive understanding of the key pathogen recognition and defense pathways to *A. lentis* in lentil. Also, the full-length gene sequences will be used in transgenic studies to further characterize function.

Mapping Populations

Mapping populations for important traits are essential genetic resources to establish trait-marker association. Therefore, efforts have been made at ICARDA and national programs to develop mapping populations for key traits in lentil (Table 6.2). RIL (recombinant inbred lines) populations were developed from the crosses made between contrasting parents for the traits of interest through single seed descent method. Indian Institute of Pulses Research (IIPR) has recently developed RIL population from a cross between ILL6002 and ILL7663 in order to identify and map early growth vigor genes in lentil. Identification of markers linked to the genes/QTLs governing these traits will help in development of genotype having high biomass at early stage. For tagging and mapping of genes of earliness, another mapping population was developed from a cross between Precoz (Medium early) and L4603 (early). CSK Himachal Pradesh Agricultural University, Palampur, India has developed recombinant inbred populations involving both intra and intersubspecific crosses that differ for rust reaction, drought tolerance, flowering time, plant vigour, shattering tolerance, seed size and seed weight.

Molecular Maps

The earlier *Lens* genetic linkage maps were constructed by using morphological and isozyme markers (Zamir and Ladizinsky 1984; Tadmor et al. 1987; Vaillancourt and Slinkard 1993). Although genetic mapping (linkage analysis) began in lentil in 1984, the first map comprising DNA based markers was produced by Havey and

Table 6.2 Mapping populations developed for various traits in lentil at ICARDA

Trait	Cross	Population size
Drought	ILL7946 × ILL7979	174
Cold	ILL4605 × ILL10657	153
Earliness	ILL7115 × ILL8009	150
Rust	ILL5888 × ILL6002	152
Fusarium wilt	ILL213 × ILL5883, Precoz × Idleb 2	150
Zn content	ILL5722 × ILL9888	177
	ILL9888 × ILL5480	149
Fe content	ILL9932 × ILL9951	193

Table 6.3 Molecular maps developed in lentil

Population used for mapping	No. of loci	Type of markers	Genetic map length (cM)	References
RILs (ILL5588×L692-16-1)	177	RAPD, AFLP, RFLP	1,073	Eujayl et al. (1998a)
F2 (ILL5588×ILL7537)	114	RAPD, ISSR	784	Rubeena et al. (2003)
Lens culinaris ssp. *Culinaris*×*L. c.* ssp. *orientalis*	161	RAPD, ISSR, AFLP, SSR	2,172	Duran et al. (2004)
RILs (ILL5588×L692-16-1)	283	SSR, AFLP	751	Hamwieh et al. (2005)
F2 (L830×ILWL77)	199	SSR, ISSR and RAPD	3843.4	Gupta et al. (2012a, b)
RIL (ILL5588×ILL5722)	196	RAPD, ISSR, EST-SSR, and SSR	1156.4	Gupta et al. (2012a, b)

Muehlbauer (1989). Subsequent maps were published by several workers (Table 6.3). With the development of PCR based markers, the number of mapped markers across the *Lens* genome increased dramatically (Kumar et al. 2011). The first extensive map comprised 177 RAPD, AFLP, RFLP and morphological markers was constructed using a RIL population from a cross between a cultivated *L. culinaris* ssp. *culinaris* cultivar and a *L. culinaris* ssp. *orientalis* accession (Eujayl et al. 1998a). Rubeena et al. (2003) published the first intraspecific lentil map comprising 114 RAPD, inter simple sequence repeat (ISSR) and resistance gene analogue (RGA) markers. Rubeena et al. (2006) reported F_2 map comprising 72 markers (38 RAPD, 30 AFLP, 3 ISSR and one morphological) spanning 412.5 cM. The first *Lens* map to include SSR markers was that of Durán et al. (2004). Hamwieh et al. (2005) added 39 SSR and 50 AFLP markers to the map constructed by Eujayl et al. (1998a) to produce a comprehensive *Lens* map comprising 283 genetic markers covering 715 cM. Subsequently, the first lentil map that contained 18 SSR and 79 cross genera ITAP gene-based markers was constructed using a F_5 RIL population developed from a cross between ILL5722 and ILL5588 (Phan et al. 2007). The map comprised seven linkage groups that varied from 80.2 to 274.6 cM in length and spanned a total of 928.4 cM. Gupta et al. (2012a) used 196 markers including new 15 *M. truncatula* EST-SSR/SSR using a population of 94 RIL produced from a cross between ILL5588 and ILL5722 and clustered into 11 linkage groups (LG) covering 1156.4 cM. An intersubspecific F_2 *Lens* linkage map consisting of 199 PCR-based markers (28 SSRs, 9 ISSRs and 162 RAPDs) mapped on to 11 linkage groups covering a distance of 3,847 cM has been constructed (Gupta et al. 2012b).

Quantitative Trait Loci (QTL) Mapping

Molecular markers linked to desirable genes/QTL have been reported for marker-assisted selection in lentil (Table 6.4). Morphological markers viz., cotyledon (*Yc*), anthocyanin in stem (*Gs*), pod indehiscence (*Pi*), seed coat pattern (*Scp*), flower

Table 6.4 Molecular markers linked to desirable genes/QTL for marker-assisted selection in lentil

Traits	QTL/genes	Type of markers	References
Ascochyta blight resistance	*QTL*	RAPD	Ford et al. (1999)
	Ra/2	RAPD, SCAR	Chowdhury et al. (2001), Taran et al. (2003)
	QTLs	AFLP	Rubeena et al. (2006)
Anthracnose resistance	*LCt-2*	AFLP, RAPD	Tullu et al. (2003), Taran et al. (2003)
Fusarium wilt resistance	*Fw*	RAPD, SSR	Eujayl et al. (1998b), Hamwieh et al. (2005)
Cold winter hardiness	*Frt*	RAPD, SSR	Eujayl et al. (1999)
		RAPD, SSR, AFLP	Kahraman et al. (2004)s
Earliness and plant height	QTL	RAPD, SSR, AFLP	Tullu et al. (2008)
Plant structure, growth habit and yield	QTL	RAPDs, ISSRs, AFLPs, SSRs	Fratini et al. (2007)
Stemphylium blight resistance	QTLs	SSRs, SRAPs, RAPDs	Saha et al. (2010a)
Rust resistance	*R*	STS, SSRs, RFPLs, RAPDs, CAPS, dCAPS	Saha et al. (2010b)
Ascochyta lentis resistance	*QTLs*	RAPD, ISSR, EST-SSR, SSR	Gupta et al. (2012a, b)

colour (*W*), radiation frost tolerance locus (*Rf*), early flowering (*Sn*) and ground colour of the seed (*Gc*) were mapped as qualitative markers because they exhibited monogenic dominant mode of inheritance (Eujayl et al. 1998a; Duran et al. 2004; Hamwieh et al. 2005; Tullu et al. 2008). Further analysis for the association between DNA markers and Fusarium wilt resistance (Fw) gene was confirmed (Eujayl et al. 1998b; Hamwieh et al. 2005). However, only SSR59-2B was closely linked with Fw at 19.7 cM (Hamwieh et al. 2005). Anthracnose disease resistance (Lct-2) was mapped by Tullu et al. (2003). To date, quantitatively inherited traits have been mapped by Duran et al. (2004) who detected five QTLs each for the height of the first ramification and flowering time, three for plant height, seven for pod dehiscence, and one each for shoot number and seed diameter. Five and four QTLs were identified for winter survival and winter injury, using a RIL population of 106 lines derived from WA8649090 × Precoz (Kahraman et al. 2004). In this study, the experiments were conducted at multiple locations and only one of five QTLs was expressed in all environments. Mapping of Ascochyta blight resistance using an F_2 population derived from ILL7537 × ILL6002 identified three QTLs accounting for 47 % (*QTL*-1 and *QTL*-2) and 10 % (*QTL*-3) of disease variation. Recently, QTLs conferring resistance to Stemphylium blight and rust diseases using RIL populations were identified in lentil (Saha et al. 2010a, b). Though the use of F_2 populations in identification of QTLs has been done widely in lentil, their use in marker-trait analysis has led to identification of only major QTLs. Thus, several minor QTLs were overlooked in such populations and identification of environmental responsive QTLs was difficult. Because quantitative traits are influenced by both genetic and environmental effects, RILs or near isogenic lines (NILs) are more suitable populations to accurately dissect their components. For ascochyta blight, three QTLs each were

detected for resistance at seedling and pod/maturity stages (Gupta et al. 2012a). Together these accounted for 34 and 61 % of the total estimated phenotypic variation and demonstrated that resistance at different growth stages is potentially conditioned by different genomic regions. The flanking markers identified may be useful for MAS and pyramiding of potentially different resistance genes into elite backgrounds that are resistant throughout the cropping season.

Application of Genomic Resources

Genomic resources can be deployed in lentil improvement programs following either molecular marker or transgenic approaches.

Determination of Molecular Diversity

Genetic diversity analysis has been studied among a set of cultivated and wild lentils using various molecular marker system and genetic materials. Earlier studies have used RFLP, AFLP and RAPD markers to assess genetic diversity and phylogenetic analyses within and among *Lens* species (Havey and Muehlbauer 1989; Abo-el-Wafa et al. 1995; Ahmad and McNeil 1996; Sharma et al. 1995, 1996; Ford et al. 1997) and gene mapping (Eujayl et al. 1998b; Tullu et al. 2003; Duran et al. 2004; Kahraman et al. 2004; Hamwieh et al. 2005). As a part of the CGIAR's Generation Challenge Program (GCP), International Center for Agricultural Research in the Dry Areas (ICARDA) has identified a composite collection of lentil germplasm and characterized them by using microsatellite markers. ICARDA holds the largest global collection of lentil with >11,000 accessions. From this collection, a global composite collection of 960 accessions (Table 6.5) representing landraces, wild relatives, elite breeding lines and cultivars was established (Furman 2006). The results indicated two major clusters separating south Asia (Nepal, India, Pakistan and Afghanistan) from the Middle East and western countries (Fig. 6.1, Hamweigh et al. in progress). The major output of this study was a reference set which represents around 15 % (135 accessions) of the global composite collection representing all the geographical regions. This set has been phenotyped for different biotic and abiotic stresses, and emerged as a useful genetic resource to start with. Recently, a set of microsatellite markers was used to study the genetic diversity of lentil mini core set. The mini core collection comprised 109 accessions from 15 countries representing 57 cultigens (including 18 breeding lines) from 8 countries and 52 wild accessions (*L. culinaris* subsp. *orientalis*, *L. culinaris* subsp. *tomentosus* and *L. culinaris* subsp. *odemensis*) from 11 countries. The total alleles detected across the microsatellite loci were 182, with a mean of 13 alleles per locus. Wild accessions were rich in allelic variation (151 alleles) compared to cultigens (114 alleles). The genetic diversity index for the microsatellite loci in the wild accessions ranged from 0.16 (SSR28 in *L. culinaris*

Table 6.5 Composition of core germplasm representing 10 % of the global lentil collection by ICARDA

Country	No. of acc.	Country	No. of acc.	Country	No. of acc.
Afghanistan	30	Germany	10	Romania	2
Albania	1	Greece	17	Russian	13
Algeria	11	Guatemala	1	Saudi Arabia	1
Argentina	6	Hungary	3	Scg	4
Armenia	3	India	192	Slovakia	1
Azerbaijan	4	Iran	103	Spain	17
Bangladesh	6	Iraq	11	Sudan	2
Belgium	1	Italy	6	Syria	70
Brazil	2	Jordan	46	Tajikistan	5
Breeding	35	Lebanon	9	Tunisia	8
Bulgaria	6	Libyan	1	Turkey	69
Canada	3	Macedonia	3	Turkmenistan	1
Chile	27	Mexico	8	Ukraine	5
China	1	Morocco	14	United States	10
Colombia	3	Nepal	28	Unknown	7
Croatia	1	Netherlands	1	Uruguay	1
Cyprus	9	Norway	1	Uzbekistan	2
Czech Republic	6	Pakistan	27	Yemen	12
Egypt	25	Pal	4	Yugoslavia	2
Ethiopia	49	Poland	4	Sum	960
France	5	Portugal	5		

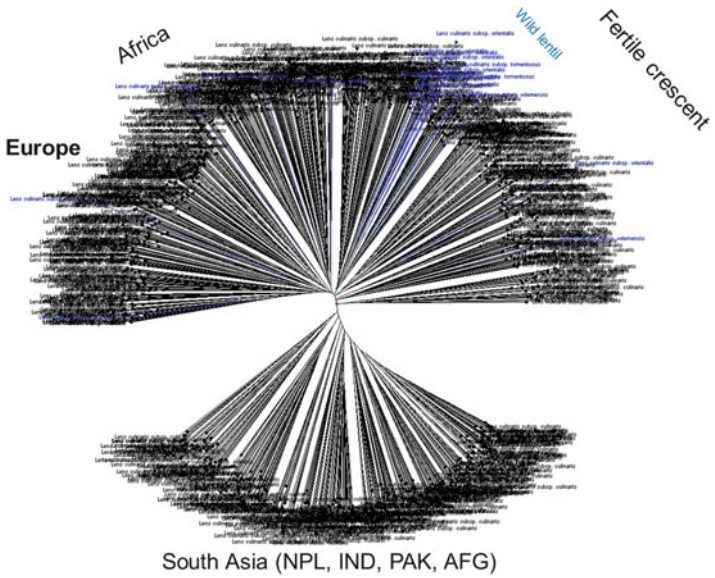


Fig. 6.1 Cluster analysis of wild and cultivated lentil accessions using 22 microsatellite markers. The results indicated two major clusters separating south Asia (Nepal, India, Pakistan and Afghanistan) from the Middle East and western countries

subsp. *odemensis*) to 0.93 (SSR66 in *L. culinaris* subsp. *orientalis*) with a mean of 0.66, while in the cultigens, genetic diversity varied between 0.03 (SSR28) and 0.87 (SSR207) with a mean of 0.65. Cluster analysis indicated two major clusters (Fig. 6.2), mainly one with the cultigens and the other with the wild accessions (Hamwieh et al. 2009).

Recently, comparative genomics approach has provided significant opportunities for analysis of genetic diversity in lentil. The conserved primers (CPs) based on *M. truncatula* EST sequences flanking one or more introns were used to sequence amplicons in 175 wild and 133 domesticated accessions. This analysis of the sequences confirmed that *L. nigricans* and *L. ervoides* are well-defined species at the DNA sequence level. *Lens culinaris* subsp. *orientalis* is the progenitor of domesticated lentil, *L. culinaris* subsp. *culinaris*, but a more specific area of origin can be suggested in southern Turkey. The study detected the divergence, following domestication, of the domesticated gene pool into overlapping large seeded (megasperma) and small-seeded (microsperma) groups and observed that lentil domestication led to a loss of genetic diversity of approximately 40 % (Alo et al. 2011).

Testing the Hybridity of F_1s

Making crosses between diverse parents is difficult in practice in lentil because of very small flowers leading to increase the chances of selfing. In addition to this, differentiating F_1 plants from selfed ones also becomes difficult due to low phenotypic diversity between the parents. Hence molecular markers have been found very useful to detect the hybridity of F_1 plants in lentil. Solanki et al. (2010) used molecular markers in lentil and detected only 21 % plants as true hybrids. These results suggest that molecular markers can reduce the time and money required to grow a population from selfed or admixed plants and increase the efficiency of plant breeders in selection of recombinant plants.

Marker Assisted Selection

Ideally, the genes controlling a trait of interest are the perfect marker for MAS. However, this is often made difficult because cloning of a gene is labor intensive and time consuming. Alternatively, marker(s) that are tightly linked to and flanking a gene locus that conditions a sizable genetic variation for the trait may be selected for with the premise that the associated chromosomal region contains the functional gene(s). Often, genetically linked markers to traits of interest are identified by coarse mapping and these have limited use in MAS because of the distance and hence chance of recombination between marker and actual gene locus. Therefore, genomic regions where the trait is mapped should be fine mapped at high resolution and be validated across genetic backgrounds in order to determine their utility in

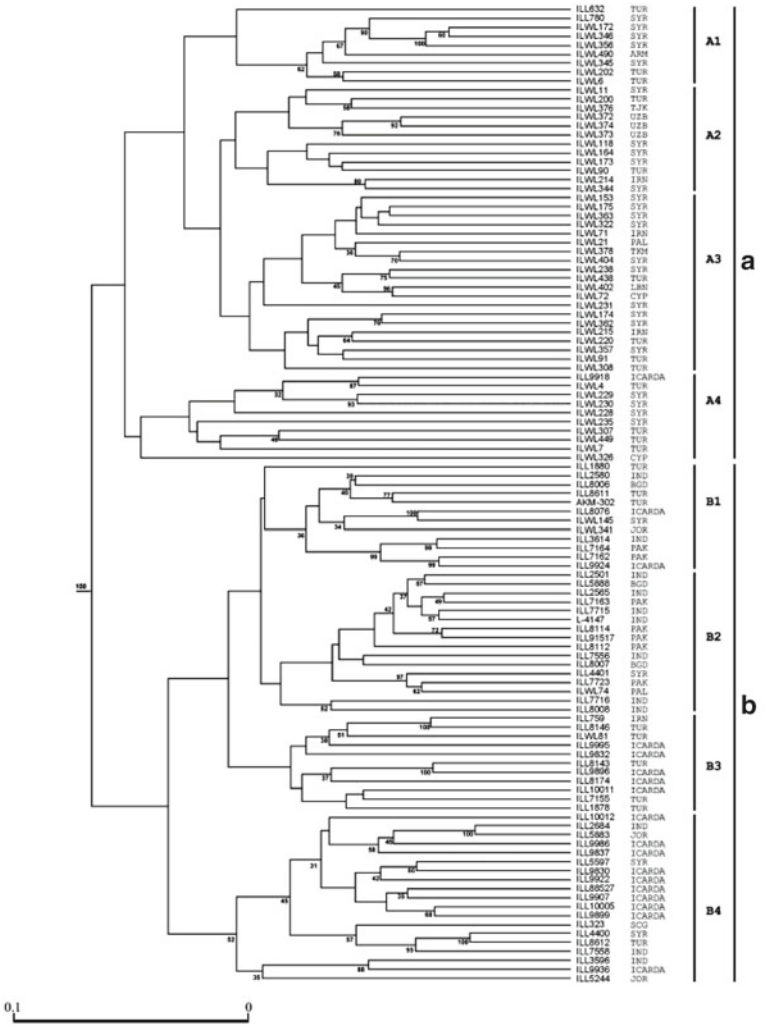


Fig. 6.2 Cluster analysis of wild and cultivated lentil accessions using 14 microsatellite markers. The groups are denoted on the *right side* as A or B, and the subgroups as A1, A2, A3, A4, B1, B2, B3, and B4. The origins of 109 lentil accessions are listed closed to the genotype numbers. Bootstrap values of above 30 % are indicated at the nodes. The abbreviations of the countries: Bangladesh (BGD), India (IND), Iran (IRN), Jordan (JOR), Pakistan (PAK), Syria (SYR), Turkey (TUR), Serbia and Montenegro (SCG), Palestine (PAL), Armenia (ARM), Cyprus (CYP), Uzbekistan (UZB), Tajikistan (TJK), Turkmenistan (TKM), Lebanon (LBN)

MAS. Also, physical characterization of genomic regions of interest will facilitate cloning of the gene to develop direct markers (candidate genes) and/or physically closer markers to the gene, increasing the reliability for MAS. The most useful marker system for MAS should be locus specific, highly reproducible and easy to discern. These include sequence tagged site (STS), sequence characterized amplified region (SCAR) or allele specific amplified primer (ASAP), specific polymorphic locus amplification test (SPLAT) and PCR-based RFLP markers. When locus specific markers are not polymorphic among the parental lines used in the breeding programs, sequence discriminative methods are required. These include SNP, cleaved amplified polymorphic site (CAPS) and derived CAPS (dCAPS) markers. Meanwhile, there are several markers available for different traits that have the potential for use in MAS and gene pyramiding. Two QTLs governing Ascochyta blight resistance were identified on LG I and II in lentil for which dominant and partial dominant gene actions were observed (Nguyen et al. 2001). These QTLs may represent the effects of the two major dominant genes previously reported for resistance in ILL7537. These include SCARW19 and SCARB18 linked to and flanking the *AbR*1 resistance loci (Nguyen et al. 2001; Taran et al. 2003). These enabled successful pyramiding of *AbR*1 and *ral*2 resistance loci together with the *LCt*2 anthracnose resistance loci (Taran et al. 2003).

Genetic Manipulation Through Transformation

Transgenic approach uses functional genes which are not available within the crossable gene pool. Thus cloned genes are important genomic resources for making genetic manipulation through transformation. Commonly, the particle bombardment and the *Agrobacterium tumefaciens* infection methods have been used to introduce genes with novel functions. With the explosion of sequence information available in the databases, transformation systems have also become useful tools to study gene function via RNA interference 'knockout', T-DNA insertion or transforming a genotype lacking a particular gene. Thus a robust, reproducible and efficient transformation system combined with a protocol to regenerate complete fertile plants from transformed cells is essential to fully study plant gene functions.

Following the initial report of shoot regeneration (Bajaj and Dhanju 1979) from apical meristems, it has been achieved routinely with different explants such as apical meristems (Bajaj and Dhanju 1979), stem nodes (Polanco et al. 1988; Singh and Raghuvanshi 1989; Ahmad et al. 1997), cotyledonary node (Warkentin and McHughen 1992; Sarker et al. 2003a), epicotyls (Williams and McHughen 1986), decapitated embryo, embryo axis and immature seeds (Polanco and Ruiz 2001) and cotyledonary petioles (Khawar and Özcan 2002). The induction of functional roots on *in vitro*-developed shoots has been the major challenge in lentil micro propagation. The difficulty to induce roots is thought to be associated with the use of cytokinin to obtain multiple shoots from the initial explants (Mohamed et al. 1992; Sarker et al. 2003b). Among the several studies conducted on root induction

from shoots, Fratini and Ruiz (2003) reported 95 % rooting efficiency from nodal segments cultured in an inverted orientation in media with 5 μM indole acetic acid (IAA) and 1 μM kinetin (KN). Sarker et al. (2003b) reported 30 % rooting efficiency on MS medium supplemented with 25 mg/l indole butyric acid (IBA). More recently Newell et al. (2006) obtained 100 % rooting efficiency on nodal micro-cuttings placed inverted in a mixture of sphagnum peat, coarse river sand and perlite at a 0.5:2:2 ratio, and concluded that the improved rooting efficiency was due to greater aeration.

To date, transformation of lentil has been reported through *A. tumefaciens-mediated* gene transfer (Warkentin and McHughen 1992; Lurquin et al. 1998; Sarker et al. 2003a) and biolistic transformation including electroporation (Chowrira et al. 1996) and particle bombardment (Gulati et al. 2002; Mahmoudian et al. 2002). Warkentin and Mc-Hughen (1992) reported the susceptibility of lentil to *A. tumefaciens* and later evaluated a number of explant types including shoot apices, epicotyl, root, cotyledons and cotyledonary nodes. All explants showed transient *b-glucuronidase* (GUS) expression at the wound sites except cotyledonary nodes, which were subsequently transformed by Sarker et al. (2003b). Öktem et al. (1999) reported the first transient and stable chimeric transgene expression on cotyledonary lentil nodes using particle bombardment. Gulati et al. (2002) reported regeneration of the first fertile transgenic lentil plants on MS medium with 4.4 μM benzyladenine (BA), 5.2 μM gibberellic acid (GA3) and chlorsulfuron (5 nM for 28 days and 2.5 nM for the rest of the culture period), followed by micrografting and transplantation in soil. The first successful work was reported by Barton et al. (1997), using pCGP1258 plasmid construct on four lentil genotypes. Khatib et al. (2007) have developed herbicide-resistant lentil through *A. tumefaciens* mediated transformation. This was achieved with the same plasmid construct pCGP1258, harbouring the *bar* gene conferring resistance to the herbicide glufosinate ammonium that was transformed using *A. tumefaciens* strain *AgL0*. Three lentil lines, ILL5582, ILL5883 and ILL5588, were used and a high selection pressure of 20 mg/l of glufosinate was applied to the explants for 18 weeks. Surviving shoots were subsequently grafted onto non-transgenic rootstock and plantlets were transferred to soil and acclimatized. The presence of the transgene was confirmed by PCR and the gene function was confirmed via herbicide application. Recently, Akcay et al. (2009) reported the production of transgenic lentil plants via *Agrobacterium*-mediated transformation and the stable transmission of the *nptII* and *gus*A genes in the subsequent generations. However, these studies were mostly confined to establish transformation techniques rather than the introduction of genes into improved varieties. Khatib et al. (2011) reported for the first time the introduction of the *dreb1a* gene into lentil for enhancing drought and salinity tolerance. The PCR results confirmed the insertion and stable inheritance of the gene of interest and *bar* marker gene in the plant genome. The Southern blot analysis revealed integration of a single copy of the transgene. The *DREB1A* gene driven by rd29A promoter transcribed in the transgenic plants by inducing salt stress in form of sodium chloride solution. The results showed that mRNA was accumulated and thus the *DREB1A* gene was expressed in the transgenic plants.

Conclusion

Pace of development of genomic resources and enabling technologies is still slow in lentil. Limited population size, low heritability, lack of candidate genes, low marker density and the difficulty in identifying beneficial alleles are the main limiting factors in genomics enabled improvement. More concerted efforts are required for developing more number of SSR and SNP markers in lentil because both are breeder friendly, highly polymorphic, evenly distributed throughout the genome and highly reproducible. The next generation sequencing technology has opened new opportunity of fast development of sequence based markers. These sequencing methodology are widely used in major legumes such as chickpea and soybean. Application of markers to practical breeding programs worldwide is still limited, and thus, more molecular maps, and genomics approaches including more gene sequences need to be developed for broadening our understanding of the complex nature of lentil genome. Genome sequencing in lentil is underway and it is expected to leverage vast genetic information to be used by lentil breeders. Lentil breeders can play an important role in development of trait specific mapping populations and precise phenotyping to establish the association of gene sequences/markers with desirable traits. Access to high-throughput genotyping and sequencing technologies is expected to speed up the genetic gain across the target environments in lentil. These developments ultimately will increase the utilization of genomic resources in genetic improvement of lentil and will lead fast track development of improved cultivars.

References

Abo-el-Wafa A, Murai K, Shimada T (1995) Intra- and inter-specific variations in *Lens* revealed by RAPD markers. Theor Appl Genet 90:335–340

Ahmad M, McNeil DL (1996) Comparison of crossability, RAPD, SDS-PAGE and morphological markers for revealing genetic relationships within and among Lens species. Theor Appl Genet 93:788–793

Ahmad M, Fautrer AG, McNeil V, Hill GD, Burritt DJ (1997) In vitro propagation of *Lens* species and their F_1 interspecific hybrids. Plant Cell Tiss Org Cult 47:169–176

Akcay UC, Mahmoudian M, Kamci H, Yucel M, Oktem HA (2009) *Agrobacterium tumefaciens*-mediated genetic transformation of a recalcitrant grain legume, lentil (*Lens culinaris* Medik). Plant Cell Rep 28:407–417

Alo F, Furman BJ, Akhunov E, Dvorak J, Gepts P (2011) Leveraging genomic resources of model species for the assessment of diversity and phylogeny in wild and domesticated lentil. J Hered 102:315–329

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

Bajaj YPS, Dhanju MS (1979) Regeneration of plants from apical meristem tips of some legumes. Curr Sci 48:906–907

Barton J, Klyne A, Tennakon D, Francis C, Hamblin J (1997) Development of a system for gene transfer to lentils. In: Proceedings of international food legume research conference III, Adelaide, South Australia, Kluwer Academic Publishers, Dordrecht, The Netherlands, p 85

Chagne D, Carlisle CM, Blond C, Volz RK, Whitworth CJ, Oraguzie NC, Crowhurst RN, Allan AC, Espley RV, Hellens RP, Gardiner SE (2007) Mapping a candidate gene (MdMYB10) for red flesh and foliage colour in apple. BMC Genomics 8:212

Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis TNH, Doyle J, Kiss GB, Young ND, Cook DR (2004) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci U S A 101:15289–15294

Choudhary S, Sethy NK, Shokeen B, Bhatia S (2008) Development of chickpea EST-SSR markers and analysis of allelic variation across related species. Theor Appl Genet 118:591–608

Chowdhury M, Andrahennadi CP, Slinkard AE, Vandenberg A (2001) RAPD and SCAR markers for resistance to ascochyta blight in lentil. Euphytica 118:331–337

Chowrira GM, Akella VF, Fuerst PE, Lurquin PF (1996) Transgenic grain legumes obtained by in planta electroporation-mediated gene transfer. Mol Biotechnol 5:85–86

Datta S, Tiwari S, Kaashyap M, Gupta PP, Choudhary PR, Kumari J, Kumar S (2011) Genetic similarity analysis of lentil (*Lens culinaris* Medik. Subsp. *Culinaris*) using cross genera legume STMS markers. Crop Sci 51:2412–2422

Duran Y, Fratini R, Garcia P, De La Vega MP (2004) An intersubspecific genetic map of *Lens*. Theor Appl Genet 108:1265–1273

Erskine W, Sarker A, Kumar S (2011) Crops that feed the world 3. Investing in lentil improvement toward a food secure world. Food Secur 3:127–139

Eujayl I, Baum M, Powell W, Erskine W, Pehu E (1998a) A genetic linkage map of lentil (*Lens* sp.) based on RAPD and AFLP markers using recombinant inbred lines. Theor Appl Genet 97:83–89

Eujayl I, Erskine W, Bayaa B, Baum M, Pehu E (1998b) Fusarium vascular wilt in lentil: inheritance and identification of DNA markers for resistance. Plant Breed 117:497–499

Eujayl I, Erskine W, Baum M, Pehu E (1999) Inheritance and linkage analysis of frost injury in lentil. Crop Sci 39:639–642

Ford R, Pang ECK, Taylor PWJ (1997) Diversity analysis and species identification in *Lens* using PCR generated markers. Euphytica 96:247–255

Ford R, Pang ECK, Taylor PWJ (1999) Genetics of resistance to ascochyta blight (*Ascochyta lentis*) of lentil and the identification of closely linked RAPD markers. Theor Appl Genet 98:93–98

Ford R, Viret L, Pang ECK, Taylor PWJ, Materne M, Mustafa B (2007) Defense responses in lentil to *Ascochyta lentis*. In: Proceedings of the Australian plant pathology conference, Australian Plant Pathology Society, Adelaide, South Australia, p 64

Fratini R, Ruiz ML (2003) A rooting procedure for lentil (*Lens culinaris* Medik.) and other hypogeous legumes (pea, chickpea and lathyrus) based on explant polarity. Plant Cell Rep 21:726–732

Fratini R, Duran Y, Garcia P, Perez de la Vega M (2007) Identification of quantitative trait loci (QTL) for plant structure, growth habit and yield in lentil. Spanish J Ag Res 5:348–356

Furman BJ (2006) Methodology to establish a composite collection: case study in lentil. Plant Genet Resour 4:2–12

Gulati A, Schryer P, McHughen A (2002) Production of fertile transgenic lentil (*Lens culinaris* Medik) plants using particle bombardment. In Vitro Cell Dev Biol 38:316–324

Gupta PK, Singh J (1981) Standard karyotype in lentil var. Plant L-639. Lentil Exp News Serv 8:23

Gupta D, Taylor PWJ, Inder P, Phan HTT, Ellwood SR, Mathur PN, Sarker A, Ford R (2012a) Integration of EST-SSR markers *of Medicago truncatula* into intraspecific linkage map of lentil and identification of QTL conferring resistance to ascochyta blight at seedling and pod stages. Mol Breed 30:429–439

Gupta M, Verma B, Kumar N, Chahota RK, Rathour R, Sharma SK, Bhatia S, Sharma TR (2012b) Construction of intersubspecific molecular genetic map of lentil based on ISSR, RAPD and SSR markers. J Genet 91(3):279–287

Hamwieh A, Udupa SM, Choumane W, Sarker A, Dreyer F, Jung C, Baum M (2005) A genetic linkage map of *Lens* sp. based on microsatellite and AFLP markers and the localization of fusarium vascular wilt resistance. Theor Appl Genet 110:669–677

Hamwieh A, Udupa SM, Sarker A, Jung C, Baum M (2009) Development of new microsatellite markers and their application in the analysis of genetic diversity in lentils. Breed Sci 59:77–86

Havey MJ, Muehlbauer FJ (1989) Linkage between restriction fragment length, isozyme and morphological markers. Theor Appl Genet 77:395–401

Kahraman A, Kusmenoglu I, Aydin N, Aydogan A, Erskine W, Muehlbauer FJ (2004) QTL mapping of winter hardness genes in lentil. Crop Sci 44:13–22

Kanazin V, Talbert H, See D, DeCamp P, Nevo E, Blake T (2002) Discovery and assay of single-nucleotide polymorphisms in barley (*Hordeum vulgare*). Plant Mol Biol 48:529–537

Kaur S, Cogan NO, Pembleton LW, Shinozuka M, Savin KW, Materne M, Forster JW (2011) Transcriptome sequencing of lentil based on second-generation technology permits large-scale unigene assembly and SSR marker discovery. BMC Genomics 12:265

Khatib F, Koudsieh S, Ghazal B, Barton JE, Tsujimoto H, Baum M (2007) Developing herbicide resistant lentil (*Lens culinaris* Medikus subsp. *culinaris*) through *Agrobacterium* mediated transformation. Arab J Plant Prot 25:185–192

Khatib F, Makris A, Yamaguchi-Shinozakic K, Kumar S, Sarker A, Erskine W, Baum M (2011) Expression of DREB1A gene in lentil (*lens culinaris* Medik. *culinaris*) transformed with the *Agrobacterium* system. Crop Pasture Sci 62:488–495

Khawar KM, Özcan S (2002) Effect of indole-3-butyric acid on in vitro root development in lentil (*Lens culinaris* Medik.). Turk J Bot 26:109–111

Kota R, Varshney RK, Thiel T, Dehmer KJ, Graner A (2001) Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.). Hereditas 135:145–151

Kota R, Rudd S, Facius A, Kolesov G, Thiel T, Zhang H, Stein N, Mayer K, Graner A (2003) Snipping polymorphisms from large EST collections in barley (*Hordeum vulgare* L.). Mol Genet Genomics 270:224–233

Kumar J, Pratap A, Solanki RK, Gupta DS, Goyal A, Chaturvedi SK, Nadarajan N, Kumar S (2011) Advances in genomics resources for improving food legume crops. J Agric Sci 150:289–318

Lurquin PF, Cai Z, Stiff CM, Fuerst P (1998) Half embryo cocultivation technique for estimating the susceptibility of pea (*Pisum sativum* L.) and lentil (*Lens culinaris* Medik) cultivars to *Agrobacterium tumefaciens*. Mol Biotechnol 9:175–179

Mahmoudian M, Yücel M, Öktem HA (2002) Transformation of lentil (*Lens culinaris* M.) cotyledonary nodes by vacuum infiltration of *Agrobacterium tumefaciens*. Plant Mol Biol Rep 20:251–257

Mohamed MF, Read PE, Coyne DP (1992) Plant regeneration from in vitro culture of embryonic axis explants in common and tepary beans. J Am Soc Hort Sci 117:332–336

Muehlbauer FJ, Cho S, Sarker A, Ford R (2006) Application of modern technologies in lentil breeding for biotic and abiotic stress resistance. Euphytica 147:149–165

Nasu S, Suzuki J, Ohta R, Hasegawa K, Yui R, Kitazawa N, Monna L, Minobe Y (2002) Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa, Oryza rufipogon*) and establishment of SNP markers. DNA Res 9:163–171

Newell C, Growns D, McComb J (2006) Aeration is more important than shoot orientation when rooting lentil (*Lens culinaris* medik.) cv. digger microcuttings In vitro. In vitro Cell Dev Biol Plant 42:197–200

Nguyen TT, Taylor PWJ, Brouwer JB, Pang ECK, Ford R (2001) A novel source of resistance in lentil (*Lens culinaris* ssp. *culinaris*) to ascochyta blight caused by *Ascochyta lentis*. Aust J Plant Pathol 30:211–215

Öktem HA, Mahmoudian M, Eyidooan F, Yücel M (1999) GUS gene delivery and expression in lentil cotyledonary nodes using particle bombardment. Lens Newsl 26:3–6

Pandian A, Ford R, Taylor PWJ (2000) Transferability of sequence tagged microsatellite site (STMS) primers across four major pulses. Plant Mol Biol Rep 18:395a–395h

Phan HTT, Ellwood SR, Hane JK, Ford R, Materne M, Oliver RP (2007) Extensive macrosynteny between *Medicago truncatula* and *Lens culinaris* ssp. *culinaris*. Theor Appl Genet 114:549–558

Polanco MC, Ruiz ML (2001) Factors that affect plant regeneration from in vitro culture of immature seeds in four lentil (*Lens culinaris* Medik.) cultivars. Plant Cell Tiss Org Cult 66:133–139

Polanco MC, Pelaez MI, Ruiz ML (1988) Factors affecting callus and shoot formation from in vitro cultures of *Lens culinaris* Medik. Plant Cell Tiss Org Cult 15:175–182

Rahman MM, Sarker A, Kumar S, Ali A, Yadav NK, Rahman ML (2009) Breeding for short season environments. In: Erskine W, Muehlbauer FJ, Sarker A, Sharma B (eds) The lentil: botany, production and use. CAB International, Wallingford, UK, pp 121–136

Reddy MRK, Rathour R, Kumar N, Katoch P, Sharma TR (2009) Cross-genera legume SSR markers for analysis of genetic diversity in Lens species. Plant Breed 129(5):514–518

Rubeena A, Ford R, Taylor PWJ (2003) Construction of an intraspecific linkage map of lentil (*Lens culinaris* ssp. *culinaris*). Theor Appl Genet 107:920–926

Rubeena A, Taylor PWJ, Ades PK, Ford R (2006) QTL mapping of resistance in lentil (*Lens culinaris* ssp. *culinaris*) to ascochyta blight (*Ascochyta lentis*). Plant Breed 125:506–512

Rudd S (2003) Expressed sequence tags: alternative or complement to whole genome sequences? Trends Plant Sci 87:321–329

Saha GC, Sarker A, Chen W, Vandemark GJ, Muehlbauer FJ (2010a) Inheritance and linkage map positions of genes conferring resistance to *Stemphylium* blight in lentil. Crop Sci 50:1831–1839

Saha GC, Sarker A, Chen W, Vandemark GJ, Muehlbauer FJ (2010b) Identification of markers associated with genes for rust resistance in *Lens culinaris* Medik. Euphytica 175:261–265

Sarker RH, Biswas A, Mustafa BM, Mahbub S, Hoque MI (2003a) Agrobacterium-mediated transformation of lentil (*Lens culinaris* Medik.). Plant Tiss Cult 13:1–12

Sarker RH, Mustafa BM, Biswas A, Mahbub S, Nahar SMM, Mahbub S, Hashem R, Hoque MI (2003b) In vitro regeneration in lentil (*Lens culinaris* Medik.). Plant Tiss Cult 13:155–163

Schmutz J, Cannon SB, Schlueter J, Jianxin M, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Dong X, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang X, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Jane GX, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the paleopolyploid soybean. Nature 463:178–183

Sharma SK, Dawson IK, Wauge R (1995) Relationship among cultivated and wild Lentil revealed by RAPD analysis. Theor Appl Genet 91:647–651

Sharma SK, Knox MR, Ellis TNH (1996) AFLP analysis of the diversity and phylogeny of *Lens* and its comparison with RAPD analysis. Theor Appl Genet 93:751–758

Simon CJ, Muehlbauer FJ (1997) Construction of a chickpea linkage map and its comparison with maps of pea and lentil. J Hered 88:115–119

Singh RK, Raghuvanshi SS (1989) Plantlet regeneration from nodal segment and shoot tip derived explants of lentil. Lens Newsl 16:33–35

Solanki RK, Singh S, Kumar J (2010) Molecular marker assisted testing of hybridity of F_1 plants in lentil. J Food Legumes 23:21–24

Somers DJ, Kirkpatrick R, Moniwa M, Walsh A (2003) Mining single nucleotide polymorphisms from hexaploid wheat ESTs. Genome 46:431–437

Tadmor Y, Zamir D, Ladizinsky G (1987) Genetic mapping of an ancient translocation in the genus *Lens*. Theor Appl Genet 73:883–892

Taran B, Buchwaldt L, Tullu A, Banniza S, Warkentin TD, Vandenberg A (2003) Using molecular markers to pyramid genes for resistance to ascochyta blight and anthracnose in lentil (*Lens culinaris* Medik.). Euphytica 134:223–230

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp *mays* L.). Proc Natl Acad Sci U S A 98:9161–9166

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422

Tullu A, Buchwaldt L, Warkentin T, Taran B, Vandenberg A (2003) Genetics of resistance to anthracnose and identification of AFLP and RAPD markers linked to the resistance gene in PI320937 germplasm of lentil (*Lens culinaris* Medikus). Theor Appl Genet 106:428–434

Tullu A, Taran B, Warkentin T, Vandenberg A (2008) Construction of an intraspecific linkage map and QTL analysis for earliness and plant height in lentil. Crop Sci 48:2254–2264

Vaillancourt RE, Slinkard AE (1993) Linkage of morphological and isozyme loci in lentil. Can J Plant Sci 73:917–926

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MTA, Azam S, Fan G, Whaley AM, Farmer AD, Sheridan J, Iwata A, Tuteja R, Penmetsa RV, Wu W, Upadhyaya HD, Yang SP, Shah T, Saxena KB, Michael T, McCombie WR, Yang B, Zhang G, Yang H, Wang J, Spillane C, Cook DR, May GD, Xu X, Jackson SA (2011) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotechnol 30:83–89

Warkentin TD, McHughen A (1992) *Agrobacterium tutumefaciens*-mediated beta-glucuronidase (GUS) gene expression in lentil (*Lens culinaris* Medik.) tissues. Plant Cell Rep 11:274–278

Weber JL (1990) Informativeness of human (dC-dA)n (dG-dT)n polymorphisms. Genomics 7:524–530

Weeden NF, Muehlbauer FJ, Ladizinsky G (1992) Extensive conservation of linkage relationships between pea and lentil genetic maps. J Hered 83:123–129

Williams DJ, McHughen A (1986) Plant regeneration of the legume *Lens culinaris* Medik (lentil) in vitro. Plant Cell Tiss Org Cult 7:149

Zamir D, Ladizinsky D (1984) Genetics of allozyme variants and linkage groups in lentil. Euphytica 33:329–336

Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003) Single nucleotide polymorphisms (SNPs) in soybean. Genetics 163:1123–1134

Zhu H, Choi HK, Cook DR, Shoemaker RC (2005) Bridging model and crop legumes through comparative genomics. Plant Physiol 137:1189–1196

Chapter 7
Advances in Cowpea Improvement and Genomics

B.B. Singh, Michael P. Timko, and Francisco J.L. Aragao

Abstract Cowpea is a diploid (2n = 2× = 22) self pollinating legume species with a genome size of 613 Mbp. Since the available genomic resources are not adequate, the use of genomics tool in cowpea breeding programme has been very limited. However, a modest beginning in developing genomic resources has led the basic foundation of use of genomic resources in cowpea improvement. In order to perform genetic and genomic analysis various markers like RFLP, RAPD, AFLP, SSR and SNP have been employed in several studies. QTLs for striga and aphid resistance have been identified and validated. However, QTLs for other agronomic traits and important diseases and pests are still to be explored. Eight linkage maps including one consensus map published so far describes a good progress in development of linkage maps. Further efforts are required to construct high-density genetic map for analyzing inheritance of target gene and localization of specific genomic regions for map based cloning. Efforts are also on sequencing of genome of this important crop. With identification of micro RNAs, ESTs, BACs and transcriptomic data sets, the cowpea genomics is gaining momentum. The need of integration of all these efforts will promote the cowpea improvement. This paper presents an overview on advances made in development of genomic resources, gene expression and regulation, marker assisted breeding and progress towards sequencing cowpea genome.

Keywords Vigna unguiculata • Molecular mapping • Marker assisted selection • Sequencing • QTLs • Genomic resources • ESTs • SNPs • Crop improvement

B.B. Singh, Ph.D. (✉)
Department of Soil and Crop Sciences, Texas A&M University, Olsen Avenue, Heep Bldg. Rm. 428, College Station, TX 77843, USA
e-mail: drbbsingh@yahoo.com

M.P. Timko, Ph.D.
Department of Biology, University of Virginia, Charlottesville, VA, USA

F.J.L. Aragao, Ph.D.
Embrapa Recursos Genéticos e Biotecnologia, Norte Brasília, DF, Brazil

S. Gupta et al. (eds.), *Legumes in the Omic Era*, DOI 10.1007/978-1-4614-8370-0_7,

Introduction

Cowpea [*Vigna unguiculata* (L.) Walp.] is a widely cultivated diploid legume species with 2n = 22 chromosomes. It is cultivated in over 65 countries covering Asia and Oceania, the Middle East, Southern Europe, Africa, southern USA and Central and South America (Singh 2005). With about 25–30 % protein in the grains and 15–18 % protein in its haulms, cowpea is a major source of dietary protein and minerals to humans as well as livestock. The average yield of cowpea is less than 500 kg/ha due to several production constraints including spreading growth habit and late maturity of traditional varieties, numerous diseases, parasitic weeds, insects and low soil fertility as well as shading due to intercropping with cereals like maize, sorghum and millet (Singh et al. 2003a, b).

Efforts have been made in the establishment of genomic resources and their application for cowpea genetic improvement. The advances began with the development and use of molecular marker technologies for diversity analysis of germplasm and in molecular breeding activities, and now include genomic scale sequence characterization, bacterial artificial chromosome (BAC) and other megabase fragment-based genomic libraries, and protocols and platforms for genome-wide gene expression profiling. This chapter briefly describes the progress made in cowpea breeding and genomics.

Genetic Resources

The world largest collection of cowpea is maintained at IITA having over 15,700 accessions of cultivated ones drawn from over 100 countries including 560 accessions of wild relatives. These have been characterized and evaluated for desirable traits and being conserved and used in the breeding program at IITA and freely made available to national breeding programs. Systematic screening of the germplasm lines has revealed extreme variation in respect of many traits such as plant pigmentation, plant type, plant height, leaf type, growth habit, photosensitivity and maturity, nitrogen fixation, fodder quality, heat and drought tolerances, root architecture, resistance to major bacterial, fungal and viral diseases, resistance to root-knot nematodes, resistance to resistance to insect pests like cowpea aphid (*Aphis craccivora*), leaf hoppers (*Empoasca signata* and *E. dolichi*), legume bud thrips (*Megalurothrips sjostedti*), pod borers (*Maruca vitrata*), pod-sucking bugs (*Clavigralla* sp.), and bruchid (*Callosobruchus maculatus*), as well as resistance to parasitic weeds (i.e., *S. gesnerioides* and *A. vogelii*), pod traits, seed traits and grain quality (Ng and Singh 1997). Based on their multiple resistances, the germplasm lines TVu 201, TVu 408, TVu 410, TVu 1190, TVu 1977 and TVu 4577 have been extensively used in the breeding program. These were also distributed to international collaborators for broad based testing. The results from these trials indicated four lines, TVu-201, TVu-1190, TVu-1977 and TVu-4577 to be resistant to many

diseases and had very high yield potential. These were described as VITA-1, VITA-3, VITA-4 and VITA-5 (Vigna IITA-1, 3, 4, and 5) respectively and subsequently released in many countries. These VITA lines were also extensively used as parents for the initial crossing programs and development of segregating populations. The focus was primarily to develop multiple disease resistant breeding lines with high yield potential. Based on the good performance across many countries, five new lines were described as VITA numbers and released in many countries. These were TVx 1193-7D as VITA-6, TVx 289-4G as VITA-7, TVx 66-2H as VITA-8, TVx 1948-01F as VITA-9, and TVx 1836-013J as VITA-10. The breeding objectives were broadened from 1980 onwards to develop a diverse set of improved cowpea varieties differing in plant type, growth habit, maturity and seed type to suit the regional preferences and cropping systems.

Breeding Progress

The global mandate for cowpea breeding has been a challenging task to the scientists at IITA because the biotic and abiotic constraints and variety requirements for cowpea differ from region to region in respect of the seed color preference, use patterns, maturity and growth habit. Thus, no single cowpea variety could be suitable for all countries and regions. Therefore, efforts have been made to develop varieties which can suit to specific circumstances.

Development of '60-Day' Erect Type Cowpea Varieties

The traditional cowpea varieties as well as the improved varieties until 1980 were medium to late maturing and semi-spreading type. Later on a need was felt for developing extra-early erect plant type cowpea varieties to be grown in areas with short rainy seasons and as a niche crop in multiple cropping systems to expand the cowpea cultivation in non-traditional areas with an yield potential between 1.5 and 2.5 t/ha within 60–65 days maturity (Singh and Sharma 1996). These were collectively called "60-day" cowpea varieties. Beginning from 1982, a large number of "60-day" cowpea varieties have been developed. Some of the prominent varieties of this group that have been released and become popular in many countries. These are IT82E-9 (black), IT 82E-60 (white blackeye), IT82E-16 (red), IT82E-18 (tan), IT82E-32 (red), IT82D-752 (tan), IT82D-789 (light brown), IT82D-889 (red), IT83S-818 (white blackeye), IT85F-867-5, IT86D-1010 (white blackeye), IT93K-452-1(white blackeye), IT97K-1042-3 (red), IT98K-1111-1(white blackeye) and IT98K-205-8 (white small eye). Of these, IT82D-889, IT83S-818, IT85F-867-5 and IT86D-1010 are resistant to over eight major viruses and IT 98K-205-8 is resistant to major viruses as well as resistant to aphid, thrips, bruchid, *Striga* and *Alectra*.

Development of Medium and Late Dual Purpose Varieties

In addition to the extra-early varieties, a number of medium maturing varieties (75–80 days) with semi-erect plant type combining multiple pest resistance and diverse seed types were also developed and distributed to national collaborators. Some of the prominent varieties of this group that have been widely tested, released and become popular in many countries are VITA-1, VITA-3, VITA-4, VITA-5, IT84S-2163, IT84S-2246-4, IT84D-449, IT84D-666, IT85F-2020, IT86D-368, IT86D-719, IT87D-697, IT87D-1627, IT88S-574-3, IT89KD-374, IT90K-277-2, IT90K-372-1-2, IT97K-368-18 and IT98K-506-1. The photosensitive late maturing cowpea varieties are commonly grown and fit well as a relay intercrop in 'millet-sorghum-cowpea', systems in many countries in West Africa and serve as dual purpose varieties providing grain as well as fodder. However, these varieties are too late and often suffer serious yield loss due to terminal drought. Therefore, selected photosensitive varieties were used as parents and a new set of improved medium-late photosensitive as well as photo-insensitive varieties which mature between 90 and 110 days were developed. Some of these combine resistance to major diseases, aphid, bruchid as well as *Striga* and *Alectra*. The promising varieties in this group that have been released many countries are IT81D-985, IT81D-994, IT89KD-245, IT89KD-288, IT89KD-391 and IT99K-216-38-1.

Vegetable Types with Bushy Growth Habit

Several countries grow the yard long cowpea varieties as a vegetable crop but these cultivars need staking to keep pods from touching the ground and rotting which involves extra cost and thus restricts the area under cultivation. Therefore, by crossing the yard long varieties with early erect types, bush-type vegetable cultivars with 30-cm long succulent pods were developed which yield up to 18 t/ha green pods with 4–6 pickings starting at 45 days after planting. Some of the promising ones are IT81D-1225-10, IT81D-1228-14, IT81D-1225-15, and IT86D-880. These cultivars have semi-erect growth habit with extra-long peduncles (40–50 cm long), protruding well over the canopy and holding the pods above the ground. Picking green pods periodically reduces the weight on peduncles and they remain upright all the time. Frequent picking also stimulates further flowering and podding on the same peduncles, which ensures a continuous supply of green pods for a 6–7 week period after the start of picking, provided soil moisture is not limiting.

Breeding for Disease Resistance

Using the resistant germplasm lines as parents in the breeding program and a combination of field and laboratory screening methods, most of the improved cowpea

varieties have been developed for combined resistance to major diseases like Cercospora, smut, rust, Septoria, scab, Ascochyta blight and bacterial blight, Macrophomina, anthracnose. Breeding for resistance to all the diseases has been easy because of simple inheritance in all the cases (Abadassi et al. 1987). Some of the best breeding lines with multiple resistances to major fungal and bacterial diseases are TVx 3236, IT81D-1228-14, IT82D-716, IT90K-277-2, IT97K-556-4, IT98K-476-8, IT97K-499-39, IT97K-1042-3, IT97K-1069-5 and IT98K-205-8. Of these TVx 3236 is a major source of resistance to scab, IT81D-1228-14 is for resistance to bacterial blight and IT97K-556-4 is for resistance to powdery mildew. Several cowpea breeding lines have also been identified with combined resistance to several major virus diseases including cowpea yellow mosaic, blackeye cowpea mosaic, southern bean mosaic, severe mosaic and many strains of cowpea aphid borne mosaic. Among these, IT82D-889, IT83S-818, IT86D-880, IT86D-1010, T90K-277-2, and IT98K-205-8 are most promising and found to be virus resistant in many countries (Van Boxtel et al. 2000) Good progress has also been made in breeding for combined resistance to several nematodes. Some of the improved breeding lines with nematode resistance are IT84S-2049, IT84S-2246-4, IT89KD-288 and IT97K-556-4 (Singh et al. 2002). Among these, IT89KD-288 is a high yielding photosensitive variety with high level of resistance to nematodes in Nigeria as well as resistant to four strains of *Meloidogyne incognita* in USA (Ehlers et al. 2000).

Breeding for Resistance to Striga *and* Alectra *Resistance*

Parasitic weeds cause considerable yield reduction in cowpea in Africa. Of these, *Striga gesnerioides* is primarily prevalent in West Africa but *Alectra vogelii* is widely distributed throughout the east, east and southern parts of Africa. Complicating the identification of *Striga*-resistant germplasm is the variable nature of the parasite with at least seven distinct races (pathotypes) of *S. gesnerioides* having now been identified. These are designated SG1 (Burkina Faso), SG2 (Mali), SG3 (Nigeria and Niger), SG4 (Benin), SG4z (Zakpota region of Benin), SG5 (Cameroon), and SG6 (Senegal) (Botanga and Timko 2006).

A local landrace, B 301 from Botswana, was found to be completely resistant to *Striga* and *Alectra* in Burkina Faso, Mali, Cameroon, Niger and Nigeria but only moderately resistant to SG4z from the Zakpota region of Benin Republic. A few other lines such as IT81D-994, IT89KD-288, 58-57 and Gorom local were found to confer complete resistance to races SG1 and SG4z from Burkina Faso and Zakpota, Benin Republic and but highly susceptible to race SG3 from Nigeria and Niger. Race-specific resistance to both parasitic weeds is inherited monogenically (Singh and Emechebe 1990; Atokple et al. 1993, 1995) and by using the complementary resistant parents in crosses, a number of new varieties have been developed with combined resistance to *Alectra* as well as all of the known races of *Striga* (Singh 2005). The most promising new cowpea varieties are IT90K-59, IT90K-76, IT90K-82-2, IT93K-693-2, T97K-499-35, and IT97K-819-118, IT98K-205-8.

Some of these lines are also resistant to bacterial blight, aphid, bruchid, thrips, and viruses with much higher yield potential than the local varieties (Singh 2005; Carsky et al. 2003). These lines also serve as a false host for *S. hermonthica* reducing its seed bank in the soil when grown as intercrop or in rotation with cereals.

Breeding for Insect Resistance

Using the available sources of resistance from germplasm lines at IITA, several improved cowpea varieties have been developed with combined resistance to aphid, thrips and bruchid (Adjadi et al. 1985; Bata et al. 1987; Singh et al. 2002). Aphid resistance is controlled by a single dominant gene which confers very high level of resistance causing death and highly reduced fecundity of aphids. Bruchid resistance is controlled by two recessive genes characterized by slow and reduced emergence of bruchids from infested seeds (Adjadi et al. 1985). This greatly minimizes seed damage due to bruchids during storage. Resistance to thrips is moderate and controlled by two recessive genes. Among several resistant breeding lines developed, IT90K-76, IT90K-59, IT 89KD-288, IT90K 277-2 and IT98K-205-8 are already popular varieties in several countries. The resistance to aphid and thrips is due to specific antibiosis and the resistance to bruchid is considered to be due to a 7s-storage protein, "vicillin" in the resistant cowpea seeds (Yunes et al. 1998). These factors are highly specific to insects only and therefore, no harmful effect to humans.

Only low level of resistance has been bred for Maruca pod borer and pod bugs. This is because none of the cultivated cowpea germplasm lines and cross-compatible wild cowpeas are resistant to Maruca pod borer. A distant wild relative of cowpea *Vigna vexillata* has shown high level of resistance to Maruca pod borer and bruchid but all the efforts made at IITA to transfer Maruca resistance genes from *Vigna vexillata* to cowpea has not been successful (Fatokun 1997). Developed through conventional breeding approaches, the new field resistant lines require only 1 or 2 sprays of insecticide for normal yield of 1.5–2.5 t compared to 5–6 sprays needed for the susceptible varieties.

Breeding for Tolerance to Drought, Heat and Cold

Since cowpea is grown in varied environments it encounters stresses such as drought, heat and cold temperatures. Also, cowpea suffers due to high temperatures in the Sahelian region. Using simple screening methods for heat and drought tolerance and root architecture, major varietal differences for all the three traits have been identified and incorporated into improved lines (Singh and Matsui 2002). Good progress has also been made at University of California, Riverside on water use efficiency, heat tolerance and chilling tolerance (Hall et al. 1997; Ismail and Hall 1998). The best drought tolerant varieties are IT89KD-374-57, IT88DM-867-11, IT98D-1399, IT98K-131-1, IT97K-568-19, IT98K-452-1, IT98K-241-2 and the best heat tolerant

lines are IT93K-452-1, IT98K-1111-1, IT93K-693-2, IT97K-472-12, IT97K-472-25, IT97K-819-43, IT and IT97K-499-38.

Breeding for Enhanced N-Fixation and Efficient Acquisition of Phosphorus

Most of the cowpeas in West Africa are grown in sandy soils which have low organic matter and low-phosphorus. Therefore, efforts are being made to screen and identify cowpea lines with enhanced nodulation and nitrogen fixation as well as efficient acquisition and utilization of phosphorus from low-P soils and rock phosphates (Sanginga et al. 2000). Recent work at IITA have shown major varietal differences in cowpea for growth, nodulation and performance under low phosphorus. Some of the promising lines under low-P condition were IT90K-372-1-2, TN5-78, IT98D-1399, TN27-80, IT99K-1060, IT89KD-374-57, TN 256-80, IT97K-1069-6 and IT98K-476-8. Screening cowpea varieties for tolerance to aluminum has also indicated major varietal differences and cowpea varieties IT91K-93-10, IT93K-2046-1 and IT90K-277-2 appear to be tolerant to aluminum and they gave higher response to phosphorus fertilization when grown in soils with aluminum toxicity problems (Kolawole et al. 2002). It is expected that the ongoing research may lead to the development of new cowpea varieties which would perform well in marginal lands where soil fertility is low.

Breeding for Improved Nutritional Traits

Following the development of a diverse set of improved cowpea varieties with high yield potential and multiple pests resistance, a systematic improvement program for nutritional and health traits was initiated in 2003. To begin with all the existing high yielding varieties and advanced breeding lines were analyzed for physical properties and protein, minerals, antioxidants and cooking properties and a great deal of variability was observed (Nielsen et al. 1993; Singh 2001). The mean values ranged from 21 to 31 % for protein, 46–79 ppm for iron, 545–1,330 ppm for calcium, 23–48 ppm for zinc, and 12,750–16,250 ppm for potassium. The best varieties in respect of high protein and high iron, zinc, calcium and potassium were IT97K-1042-3 and IT98K-205-8. The IT97K-1042-3 was also best for antioxidant activity.

Genome and Genome Size

The size of the cowpea genome was initially estimated at 613 Mbp (Arumuganathan and Earle 1991) and more recently at 620 Mb (Varshney et al. 2009) making it one of the smallest among the legumes and at the lower end of plant genomes in general.

Initially, efforts aimed at developing genomic resources were hampered owing greatly to the fact that cowpea was an orphan crop with little socioeconomic importance in the developed world. Gradually this has changed with the recognition of the broader importance of the crop. Among the first attempts at characterizing the gene content and complexity of the cowpea genome was the work of Timko et al. (2008) who applied a reduced representational approach known as methylation filtration (MF) to overcome the presence of ubiquitous repetitive DNA and capture only the hypomethylated, gene-rich coding sequences in the genome of the African cowpea cultivar IT97K-499-35. Using MF these investigators were able to achieve a 4.1-fold enrichment for the gene-rich space of cowpea and generated 263,425 gene-space sequence reads (GSRs) that could be assembled into 41,260 unigenes representing 19,786 unique GenBank accession numbers (Chen et al. 2007). Additional information on the cowpea genespace can be found on the Cowpea Genomics Knowledge Base (CGKB) website (http://cowpeagenomics.med.virginia.edu/CGKB/index.pl). The CGKB website provides an annotated, well-organized, and rigorously analyzed dataset of sequences as a resource for cowpea researchers and pan-legume crop specialists including a list of over 1,000 predicted and confirmed simple sequence repeat (SSR) primer combinations that can and in some cases have already been used for diversity analysis and molecular mapping. Additional SSR primer combinations based on expressed sequence tags (EST) (Gupta and Gopalakrishna 2010) and GSR sequences (Xu et al. 2010) can also be found in the literature.

Genomic Resources

Molecular Markers

The recent marker repertoire has enhanced our understanding of cowpea's genome structure and organization. Several markers like RAPD, SSR, AFLP and ISSR have been used to reveal the genetic diversity in cowpea. RAPD technology was proved to be a useful tool in the characterization of the genetic diversity among cowpea cultivars by Zannou et al. (2008) Malviya et al. 2012; Nkongolo 2003 and Chen et al. 2008. SSR is the most frequently used marker in the genetic diversity analysis of cowpea.

The earliest cowpea SSR research was conducted by Li et al. (2001) by developing 27 SSR primers. After that, SSR research on cowpea for assessing genetic diversity from different areas, mainly Africa and Asia, has been carried out. Africa is the diversity center of wild cowpea, which was proved by Ogunkanmi et al. (2008) with SSR analysis. Asare et al. (2010) utilized SSR molecular marker to evaluate genetic diversity and phylogenetic relationships among 141 cowpea accessions collected throughout the nine geographical regions of Ghana. Badiane et al. (2012) assessed the genetic diversity and phylogenetic relationships among 22 local cowpea varieties and inbred lines collected throughout Senegal by SSR markers and developed a set

of 44 polymorphic primer combinations from cowpea genomic or expressed sequence tags. Sawadogo et al. (2010) evaluated the genetic diversity and phylogenetic relationships among cowpea genotypes used in breeding for resistance to *Striga gesnerioides* in Burkina Faso using simple SSR molecular markers. Very few primer combinations showed polymorphic bands capable of discriminating *Striga*-resistant from susceptible cultivars, which revealed a high efficiency of SSR markers. Lee et al. (2009) estimated the genetic diversity of 492 Korean cowpea landrace accessions using six SSR markers. Xu et al. (2010) assessed the genetic diversity of asparagus bean cultivars from different geographical origins in China by EST-derived and GSS-derived SSR markers.

AFLP is recognized as one of the most efficient molecular markers. Coulibaly et al. (2002) employed AFLP to evaluate genetic relationships within a total of 117 cowpea accessions to assess the organization of their genetic diversity. Fang et al. (2007) examined genetic relationships among 60 advanced breeding lines from six breeding programs in West Africa and USA and 27 landrace accessions from Africa, Asia and South America. AFLP markers with six near infrared fluorescence labeled *Eco*RI + 3/1bases/*MseI* + 3/1bases primers sets were used in the study. Tantasawat et al. (2010) estimated genetic diversity and relatedness of 23 yardlong bean (*Vigna unguiculata* spp. *sesquipedalis*) accessions and seven accessions of a hybrid between cowpea (*V. unguiculata* spp. *unguiculata*) and dwarf yardlong bean in Thailand by morphological characters, SSR and ISSR markers

Genetic Maps

The first attempts at linkage mapping used a variety of tools aimed at detecting molecular polymorphisms such as restriction fragment length polymorphism (RFLP) analysis, randomly amplified polymorphic DNA (RAPD) detection, etc., and were successful in providing a baseline for more detailed genomic analyses. Since the first cowpea genetic mapping attempts (Fatokun et al. 1993; Menéndez et al. 1997), there has been a progression of increasingly informative maps with a greater number of traits analyzed and greater depth of marker coverage (Table 7.1).

Ouédraogo et al. (2002) offered the most comprehensive coverage, integrating amplified fragment length polymorphism (AFLP), RFLP, and RAPDs markers with numerous phenotypic characteristics and biochemical traits, into 11 linkage groups (LGs) spanning a total of 2,670 cM, with an average distance of 6.43 cM between markers. The use of this genetic map and its derivatives allowed the development of effective molecular markers for use in marker assisted breeding and selection strategies aimed at incorporating resistance to various biotic constraints into local germplasm (Timko et al. 2007). A genetic linkage map based on segregation of simple sequence repeat (SSR) markers has recently been developed using a recombinant inbred (RI) population of 159 individuals derived from a cross between the breeding line 524B, a California Blackeye type, and 219-01, a perennial wild cowpea from Kenya (Andargie et al. 2011). This genetic map contains approximately 202

Table 7.1 Molecular maps developed in cowpea

Mapping population	Parents	Markers	Number of markers	Average distance (cM)	Genetic length (cM)	Linkage groups	Reference
F2	IT 84S-2246-4×NI 963	RFLP		7.70	680	11	Young (1999)
RILs	IT84S-2049×524B	AFLP, RAPD, RFLP	242	6.43	2,670	11	Ouédraogo et al. (2002)
RILs	Six pairs of parents	SNP	928	0.73	680	11	Muchero et al. (2009a)
RILs	524B×219-01	SSR	639	3.00	677	11	Andargie et al. (2011)
RILs	JP81610×TVnu457	SSR		3.96	852.4	11	Kongjaimum et al. (2012)
RIL	Zhijiang282×ZN016	SNP, SSR		1.98	745	11	Xu et al. (2011)
RIL	IT84S-2049×524B	RAPD, RFLP, AFLP		6.40	972	12	Menéndez et al. (1997)
RIL	13 pairs of crosses	SNP	1,107	Not given	680	11	Lucas et al. (2011)

markers placed in 11 LGs spanning 677 cM, with an average distance between markers of 3 cM. Since the cross involved both a domesticated and wild forms of cowpea, the investigators were able to map agronomic traits related to domestication such as seed weight and pod shattering, as well as floral characteristics.

The advent of new and improved technologies brought rapid and significant refinements in the cowpea genetic map. Among these technologies was the development of platforms for high throughput DNA and cDNA sequencing and single nucleotide polymorphism (SNP) detection. Using SNP assays, Muchero et al. (2009a) were able to map 928 expressed sequence tag (EST)-derived SNPs using an Illumina 1536 GoldenGate platform. This map represented a substantial improvement over previously available genetic maps (Menéndez et al. 1997; Ouédraogo et al. 2002) because it was not population specific and surveyed polymorphism at 1,536 identical loci in six recombinant inbred line (RIL) populations. Building upon this work, a new consensus map containing 1,107 EST-derived SNP markers (856 bins) has been recently reported by Lucas et al. (2011). This new map was developed by integrating 13 population-specific maps and contains 1,107 markers. It is noteworthy that not only these investigators were able to add 179 new markers, an almost 20 % increase in marker density compared to the earlier consensus map created by Muchero et al. (2009a), but the number of informative positions increased on an average by 19 bins per linkage group and the average distance between informative positions was reduced from 1.05 to 0.79 cM. The consensus genetic linkage map for cowpea is given in Fig. 7.1. The SNP-based maps have an additional value as the polymorphic loci mapped are associated with expressed genes (see Lucas et al. 2011). As a consequence of being associated with a known coding region rather that a random or repetitive sequence, it is possible to examine synteny of these gene positions among closely and more distally related legume species. The use of SNPs in syntenic comparisons both with closely related species and subspecies (such as *V. unguiculata* subsp. *sesquipedalis*) and more distally related genera such as *Glycine*, *Medicago*, and *Phaseolus* has been described by Varshney et al. (2009), Lucas et al. (2011).

QTLs

Mapping more QTLs of quantitative trait by analyzing the linkage between molecular marker and those traits are significant. Kongjaimum et al. (2012) identified one major and six minor QTLs for pod length. Andargie et al. (2011) identified the QTLs for agronomic traits related to domestication (seed weight, pod shattering) by SSR markers. Six QTL for seed size were revealed with the phenotypic variation ranging from 8.9 to 19.1 g/100 seeds. Four QTLs for pod shattering were identified with the phenotypic variation ranging from 6.4 to 17.2 %. The QTL for seed size and pod shattering mainly clustered in two areas of LGs 1 and 10. Fatokun et al. (1992) identified major QTLs for seed weight. Muchero et al. (2009b)

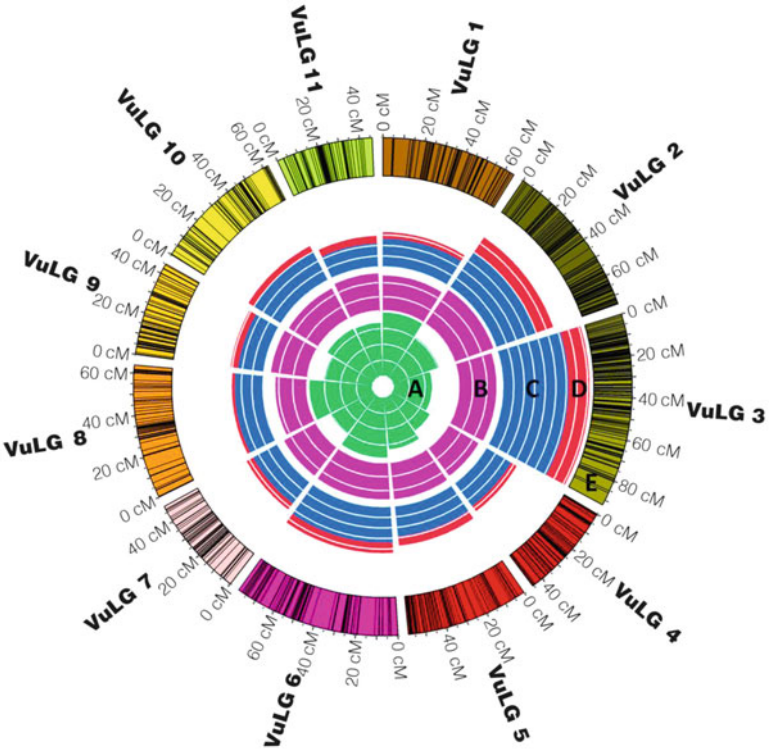


Fig. 7.1 Consensus genetic map of cowpea and parameters depicting map characteristics. (**a**) Average distance between bins (0.25 cM). (**b**) Average number of markers per bin (0.5 units). (**c**) Number of bins (25 units). (**d**) Number of markers (25 units). (**e**) Bin locations. (**c**) and (**d**) begin at the same radial position [Reprinted from Lucas MR, Diop NN, Wanamaker S, Ehlers JD, Roberts PA and Close TJ (2011) Cowpea–soybean synteny clarified through an improved genetic map. Plant Genome 4: 218–225. With permission from ACSESS-Alliance of Crop, Soil, and Environmental Science Societies. Copyright 2011 © Crop Science Society of America]

reported the mapping of 12 QTL associated with seedling drought tolerance and maturity. Regions harboring drought-related QTL were observed on linkage groups 1, 2, 3, 4, 6, 7, 9 and 10 accounting for between 4.7 and 24.2 % of the phenotypic variance. Further, two QTLs for maturity were mapped on linkage groups 7 and 8 separately from drought-related QTL.

A few QTLs of resistance to disease and insects have also been identified. For cowpea bacterial blight, Agbicodo et al. (2010) identified three QTLs, CoBB-1, CoBB-2 and CoBB-3 on linkage group LG3, LG5 and LG9, respectively. Besides, Muchero et al. (2011) identified the QTL for *Macrophomina phaseolina* resistance and maturity. Muchero et al. (2010) also identified three QTL for resistance to *Thrips tabaci* and *Frankliniella schultzei* based on an AFLP genetic linkage map.

ESTs

Among the areas where remarkable progress has been made in recent years is the significant expansion in the number a genomic and transcriptomic sequences (i.e., cDNA, expressed sequence tags, etc.) of cowpea origin available in public databases. Multiple cDNA libraries and approximately 190,000 cDNA sequences and 189,779 ESTs are publicly available from GenBank at NCBI. These represent various cowpea genotypes with the greatest proportion coming from sequence projects carried out by researchers at the University of California, and Department of Energy Joint Genome Institute, USA. For researchers, the cowpea EST assemblies are available through the HarvEST: Cowpea website (http://harvest.ucr.edu, HarvEST: Cowpea 1.27)

BAC Libraries

At least three different bacterial artificial chromosome (BAC) libraries have been produced for cowpea in recent years. The first was created from the IITA advanced breeding line IT97K-499-35 at the University of Virginia (now available through Amplicon Express, http://ampliconexpress.com/aexPremadeLib.php) and representing approximately 6× coverage of the cowpea genome. A 10× library has been constructed by George Bruening and Doug Cook (University of California, Davis; Varshney et al. 2009) from cowpea cultivar Blackeye 5 and used to generate approximately 36.7 Mbp of BAC end sequence (BES). Lastly, a second library from IT97K-499-35, consisting of approximately 60,000 BAC clones (yielding 17× genome coverage) was produced by Tim Close, Jeff Ehlers and Phil Roberts (University of California, Riverside). This library was subjected to automated, high-throughput, high-information-content fingerprinting (Luo et al. 2003) allowing Mingcheng Luo (University of California, Davis) and his colleagues to assemble a physical map of the cowpea genome. The current physical map is an assembly of 43,717 BACs with a depth of 11× genome coverage.

Small RNAs

Among the more recent developments impacting our understanding of the factors that control gene expression was the discovery of small non-coding RNAs in plants (sRNAs). There are two main types of sRNAs based on their biogenesis: microRNAs (miRNAs) and small interfering RNAs (siRNAs). miRNAs are 20–24 nucleotides long and generated by one of the Dicer-like (DCL) proteins from RNA precursors that fold into stem-loop structures MiRNAs regulate gene expression by directing mRNA cleavage or translational repression and have now been shown to be involved

in a variety of developmental processes and responses to various abiotic and biotic stresses (Jones-Rhoades et al. 2006; Sunkar et al. 2007; Brodersen et al. 2008). Several reports have appeared in the literature that examine miRNAs in cowpea. Lu and Yang (2010) used an *in silico* approach to identify 47 potential miRNAs in cowpea belonging to 13 miRNA families previously identified in other plant species. Among these, there were about 30 miRNAs predicted to target genes encoding transcription factors or enzymes participating in the regulation of development, growth, metabolism, and other physiological processes. In another study, 18 conserved miRNAs belonging to 16 families were identified. Paul et al. (2011) similarly used a comparative genomic approach and were able to identify 18 conserved *V. unguiculata* miRNAs belonging to 16 distinct miRNA families. Fifteen of the potential miRNAs were predicted to target transcription factors, and the investigators were able to experimentally validate seven of them as being present and up-regulated in roots during salt stress. In a related study, Barrera-Figueroa et al. (2011) used a combination of NextGen sequencing of sRNA and comparative bioinformatics to identify miRNAs in cowpea specifically associated with drought tolerance. These investigators were able to identify 157 miRNA genes that belong to 89 families including 44 which they were able to predict as drought-associated miRNAs. In addition, about 30 were up-regulated in drought condition and 14 were down-regulated. Many of the targets identified for these miRNAs were transcription factors associated with drought and other stress responses in cowpea.

Transcriptomic Data-Sets

A few transcriptomic datasets have been developed in cowpea indicating that many genes are expressed during drought, extreme temperature, nitrogen deficient conditions as well as during symbiosis and iron accumulation. Several transcripts known as CPRD (cowpea clones responsive to dehydration), CPRD 8, CPRD 14, CPRD 22 and VuNCED1 encode a 9-cisepoxycaratenoid dioxygenase responsible for abscisic acid (ABA) biosynthesis during drought, high salinity and heat stresses that are highly expressed (Iuchi et al. 1996, 2000). Recently, uncharacterized genes which are down-regulated in drought conditions were reported by Coetzer et al. (2010) by using suppression subtractive hybridization. Membrane stability and membrane lipids play greater role in tolerance against drought. Cystatin and aspartic protease are two important proteins related to membrane stability. The transcripts coding these proteins VuC1 and VuAP1 were isolated in drought tolerant cowpea cultivars subjected to water deficit and their expression localized in different organs (de Carvalho et al. 2001; Diop et al. 2004). The investigators reported that, the expression of the gene encoding phospholipase D1 (*VuPLD1*) was moderately increased in the drought tolerant cowpea cultivars (Maarouf et al. 1999), in that phospholipase D is a major lipid degrading enzyme in plants sensitive to drought. In heat stress conditions, analysis of transcripts expression showed 600 bands, among which 55 and 9 were up-regulated and repressed, respectively (Simoes-Araujo et al. 2002).

In other conditions such as in nitrogen deficiency, a decrease of *pur*5 transcript level which codes aminoimidazole ribonucleotide synthetase involved in purine synthesis. In symbiotic association with *Rhizobium*, the gene encoding for leghaemoglobin (*lbll*), a gene similar to the soybean leghaemoglobin *lbll* was found abundantly expressed in cowpea. These transcripts are useful resources for cowpea improvements.

Gene Expression Patterns and Their Regulation

To date only limited information is available on global transcription changes in cowpea plants during developmental and under normal physiological and aphysiological conditions such as biotic and abiotic stress conditions. Using the ~43,253 annotated unigenes obtained from sequencing of the MF gene space from cowpea a 385,000 feature long oligonucleotide-microarray (Roche–NimbleGen) was designed that represents each predicted gene coding sequence with 3–6 long oligos (60-mers) (Huang K, Mellor KE, and Timko, MP, unpublished data). This microarray was then used to examine global changes in gene expression in the roots of the cowpea cultivar B301 during compatible (susceptible) and incompatible (resistant) interactions with *S. gesnerioides* races SG4z and SG3 at 6 days and 13 days post-inoculation (dpi), early and late stages of the resistance response, respectively (Huang K, Mellor KE, and Timko, MP, unpublished data). A total of 111 genes were differentially expressed in B301 roots at 6 dpi, with this number increasing to 2,102 genes at 13 dpi. At 13 dpi during compatible (susceptible) interactions of B301 with SG4z a total of 1,944 genes were differentially expressed. Genes and pathways involved in signal transduction, programmed cell death and apoptosis, and defense response to biotic and abiotic stress were differentially expressed in the early resistance response, whereas at the latter time point enrichment was primarily for defense related gene expression, and genes encoding components of lignifications and secondary wall formation. In compatible interactions (B301–SG4z), multiple defense pathways were repressed including those involved in lignin biosynthesis, secondary cell wall modifications, while cellular transport process for nitrogen and sulfur were increased. These studies show that distinct changes in global gene expression profiles occur in host roots following successful and unsuccessful parasitism attempted by *Striga*. Induction of specific defense related genes and pathways define components of a unique resistance mechanism. Some genes and pathways up-regulated in the host resistance response to SG3 are repressed in the susceptible interactions suggesting that the parasite is targeting specific components of host defense.

Prior to the availability of a cowpea microarray platform, Das et al. (2008) were able to demonstrate that the Affymetrix soybean genome array is a satisfactory system for identification of single feature polymorphisms (SFPs) useful in the development of molecular markers for genetic mapping. Subsequently, the use of this heterologous platform was also shown to be useful in global gene expression

analysis. In order to elucidate cowpea response to root-knot nematodes, Das et al. (2010) examined the transcriptional changes in roots of resistant genotype CB46 and a susceptible near-isogenic lines (null-*Rk*) following infection with *Meloidogyne incognita* using a soybean Affymetrix GeneChip expression array. These investigators found that at 3 days post-inoculation (dpi) 746 genes were differentially expressed in incompatible interactions (infected resistant tissue compared with non-infected resistant tissue) and 623 genes were differentially expressed in compatible interactions (infected susceptible tissue compared with non-infected susceptible tissue). At later stages of nematode infection (i.e., 9 dpi) 552 genes were differentially expressed in incompatible interactions and 1,060 genes were differentially expressed in compatible interactions.

Using a different approach for monitoring global changes in gene expression, Coetzer et al. (2010) recently examined differential gene expression in drought stressed and unstressed cowpea plants by comparing the effects of water deprivation on drought tolerant (IT96D-602) and drought susceptible (Tvu7778) breeding lines developed at the International Institute of Tropical Agriculture (IITA). These investigators used suppression subtractive hybridization (SSH) to create forward and reverse cDNA libraries enriched for cowpea drought response genes. They then selected clones for sequence characterization and quantitative reverse transcription PCR based on the calculation of enrichment ratios using a statistical software pipeline they developed for the analysis (SSH screen 2.0.1; available from http://microarray.up.ac.za/SSHscreen). From the analysis they were able to identify a set of clones representing drought-induced cowpea genes as well as a group of genes significantly down-regulated by the drought stress genes. Among up-regulated category, genes were encoding a late embryogenesis abundant Lea5 protein, a glutathione S-transferase, a thaumatin, a universal stress protein, and a wound induced protein. Among the down-regulated category a lipid transfer protein and several components of photosynthesis were identified.

Marker Assisted Breeding

Marker assisted breeding was successfully employed in developing cowpea cultivars resistant to a parasitic weed, Striga gesnerioides. The SCAR and other PCR amplifiable markers were found capable of tracking most of the major race specific resistant genes to *S. gesnerioides* in West Africa (Boukar et al. 2004; Timko et al. 2007; Li et al. 2009) and the subsequent exploitation of one of these marker SSR 1 facilitated the positional cloning and characterization of the nuclear genes conferring resistance to the noxious pest (Li and Timko 2009). Besides *Striga gesnerioides*, markers were also found to be associated with the rust caused by *Uromyces vignae* An AFLP marker (E-AAG/M-CTG) was converted to a SCAR marker, named ABRSAAG/CTG 98, and the genetic distance between the marker and Rr1 gene was estimated to be 5.4 cM (Li et al. 2007). Myers et al. (1996) found one RFLP marker, bg4D9b, to be tightly linked to the aphid resistance gene (Rac 1).

The close association of rac1 and RFLP bg4D9b presented real potential for cloning this insect resistant gene. Inspite of such progress made, more concerted efforts are required to accelerate in marker assisted breeding to develop high yielding and disease resistant cultivars in cowpea.

Sequencing of Cowpea Genome

The first attempted full genome sequence was recently reported by Close et al. (2011). In this study genomic DNA from IT97K-499-35 (an improved breeding line combining genes for resistance to many diseases, insects and Striga) was shotgun sequenced using an Illumina GAII sequencer with TrueSeq chemistry in a paired-end format. The Illumina sequences (296,868 contigs with total length of ~186 MB, available at http://www.harvest-blast.org) were then assembled using SOAP denovo together with a combination of 260,642 cowpea gene-space random shotgun sequences (Timko et al. 2008) and 30,527 BAC end sequences (obtained from M.-C. Luo, UC Davis, http://phymap.ucdavis.edu:8080/cowpea), 54,123 cowpea Genome Survey Sequences (GSS) from dbGSS of GenBank http://and cowpea EST assembly to yield a draft cowpea genome assembly.

Genetic Transformations

The transformation systems developed have been used to introduce genes related to important agronomic traits into cowpea. The first report on the regeneration and stable transformation of cowpea expressing a gene of agronomic importance appeared in 2008 when a transgenic line that expressed some degree of insect resistance was generated (Solleti et al. 2008a, b). The establishment of an *Agrobacterium tumefaciens*-mediated transformation protocol using geneticin and supplementation of post-selection media with BA (Solleti et al. 2008b). The strategy was based on the use of the gene for alpha-amylase inhibiting protein (*aAI-1*) from common bean (*Phaseolus vulgaris*) as a means of conferring resistance against different insects. The efficiency of transformation in this case was enhanced by using multiple copies of the gene *vir*, co-culture of explants in the presence of thiol compounds and by sequential selection using geneticin (Solleti et al. 2008a). The work reported up to 82.3 % decrease in insect susceptibility in transgenic plants when exposed to pulse beetle (*Callosobruchus chinensis*) (Solleti et al. 2008a). This successful demonstration of cowpea resistance using *aAI-1* gene was followed by the report of another considerable resistance against *Maruca vitrata* by T_3 progenies after transformation of nodal cuttings with a plasmid harboring *Cry1Ab*, the now popular gene for protein toxin from *Bacillus thurigiensis*, using *nptII* as a selectable marker under the control of 35S of CaMV (Adesoye et al. 2008, 2010). These transgenic lines were generated by the T. J. Higgins's group at CSIRO (Australia).

A number of field trials have been going on in the last couple of years. Cowpea plants with high degree of resistance against *Maruca vitrata* and *Callosobruchus maculatus* have been subjected to field trials to test agronomic performance and insect resistance in Puerto Rico and Nigeria with promising results (T. J. Higgins, CSIRO personal communication/http://www.csiro.au/people/TJ.Higgins.html).

We have explored the interfering RNA (RNAi) mechanism to generate transgenic lines that are simultaneously resistant to the *Cowpea severe mosaic virus* (CPSMV) and *Cowpea aphid-borne mosaic virus* (CABMV) (data not published). In addition, plants are also extremely tolerant (more than three times the recommended commercial dose) to herbicides from imidazoline class. Another important candidate gene of great potential in improving cowpea is cystatin, a cysteine proteinase inhibitor with potential as a pest resistance conferring agent. We are currently trying to develop transgenic cowpea expressing chicken cystatin with a view to expressing insecticidal activity against bruchids.

Although cowpea is an important source of nutrients, including several amino acids, it is deficient in sulfur-containing amino acids, a trait common in most legumes. Several strategies have been devised to address this using transgenic technology in a number of legumes. Our group is using a transgenic approach to introduce methionine-rich protein in cowpea using the gene for δ-zein from maize.

In the last few years, significant progress has been made to establish different protocols and their application in the development of transgenic cowpea. There have been important findings that started with obtaining transgenic callus; from that came transgenic plants that exhibited mendelian segregation, culminating in recent findings that have led to the production of transgenic cowpea with agronomic traits. Currently, various research groups in countries including Australia, Brazil, India and Nigeria possess transformation systems that can be used to obtain useful genetically modified lines. Nevertheless, in our experience, only one out of 20 independent transgenic lines obtained has the potential to be introduced into a breeding program to generate a commercial variety. Consequently, despite of having suitable cowpea transformation systems, these technologies should be improved to accelerate the development of cowpea varieties with improved agricultural characteristics.

Conclusion and Perspectives

With the modest beginning, cowpea genomics is now progressing at a rapid pace. Molecular markers are essential resources for accelerating the breeding efforts for cowpea improvement. However, studies of molecular markers on cowpea are meager in comparison to other legumes like soybean and common bean. Therefore, it is necessary to make more serious research efforts in identification of molecular markers for cowpea breeding. Further a very few molecular markers have been found which were linked to resistance gene. There are a lot of diseases like rust, powdery mildew, fusarium wilt, and insect pests like bean weevil and pod borer for which there is need to identify more molecular markers linked to these disease and insect

resistance genes. QTL studies of quantitative traits in the crop are few, which need to be accelerated further for many important agronomic and economic characters such as yield, protein content, and maturity. The QTLs so identified would be useful for research on marker assisted breeding, mechanism of heterosis, genetic diversity, isolation and cloning of gene (s) associated with quantitative trait. The progress towards cowpea genome sequencing (Timko et al. 2008) in combination with the availability of genomic resources from other model legumes would help identify candidate genes that govern the agronomically important traits. After finalization of sequencing and the annotation of genome more efforts need to be done to understand the interactions between the small non coding RNA (small interfering RNA, micro RNA, trans-acting RNA, etc.) (Borsani et al. 2005). There is need for more studies to be done on cowpea proteome, metabolome, lipidome and ionome analyses. All these efforts are needed to complement to improve cowpea for higher production, resistance against key pests and diseases and quality.

Acknowledgements M.P.T. was supported by grants from the National Science Foundation (DBI-0701748 and IBN-0322420) and the Kirkhouse Trust.

References

Abadassi JA, Singh BB, Ladiende TAO, Shoyinka SA, Emechebe AM (1987) Inheritance of resistance to brown blotch, septoria leaf spot and scab in cowpea (*Vigna vexillata*). Indian J Genet 47:299–303

Adjadi O, Singh BB, Singh SR (1985) Inheritance of bruchid resistance in cowpea. Crop Sci 25:740–742

Adesoye A, Machuka J, Togun A (2008) *CRY 1AB* transgenic cowpea obtained by nodal electroporation. Afr J Biotechnol 7:3200–3210

Adesoye A, Machuka J, Togun A (2010) Transformation of cowpea (*Vigna unguiculata* L. Walp.) by *Agrobacterium* infiltration. J Appl Biosci 30:1845–1860

Agbicodo EM, Ftokun CA, Bandyopadhyay R, Wydra K, Diop NN, Muchero W (2010) Identification of markers associated with bacterial blight resistance loci in cowpea. Euphytica 175:215–226

Andargie M, Pasquet RS, Gowda BS, Muluvi GM, Timko MP (2011) Construction of a SSR-based genetic map and identification of QTLs for yield and domestication traits using recombinant inbred lines from a cross between wild X cultivated cowpea (*V. unguiculata* (L.) Walp.). Mol Breed 28:413–420

Arumuganathan K, Earle E (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

Asare AT, Gowda BS, Galyuon KA, Abogaye LL, Takramma JF, Timko MP (2010) Assessment of genetic diversity in cowpea (*Vigna unguiculata*) germplasm from Ghana using simple sequence repeat markers. Plant Genet Res 8:142–150

Atokple IDK, Singh BB, Emechebe AM (1993) Independent inheritance of *Striga* and *Alectra* resistance in cowpea genotype B 301. Crop Sci 33:714–715

Atokple IDK, Singh BB, Emechebe AM (1995) Genetics of resistance to *Striga* and *Alectra* in cowpea. J Hered 86:45–49

Badiane FA, Gowda BS, Cisse N, Diouf D, Sadio O, Timko MP (2012) Genetic relationship of cowpea (*Vigna unguiculata*) varieties from Senegal based on SSR markers. Genet Mol Res 11:292–304

Barrera- Figueroa BE, Gao L, Diop NN, Wu Z, Ehlers JD, Roberts PA, Close TJ, Zhu JK, Liu R (2011) Identification and comparative analysis of drought associated micro RNA in two cowpea genotypes. BMC Plant Biol 11:127. doi:10.1186/1471-2229-11-127

Bata HD, Singh BB, Singh SR, Landeinde TAO (1987) Inheritance of resistance to aphid in cowpea. Crop Sci 27:892–894

Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulated salt tolerance in *Arabidopsis*. Cell 123:1279–1291

Botanga CJ, Timko MP (2006) Phenetic relationships among different races of *Striga gesnerioides* (Willd.) Vatke from West Africa. Genome 49:1351–1365

Boukar O, Kong L, Singh BB, Murdock L, Ohm HW (2004) AFLP and AFLP-derived SCAR markers associated with *Striga gesnerioides* resistance in cowpea (*Vigna unguiculata* (L.) Walp.). Crop Sci 44:1259–1264

Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, Sieburth L, Voinnet O (2008) Wide spread translational inhibition by plant miRNAs and siRNAs. Science 320:1185–1190

Carsky RJ, Akakpo C, Singh BB, Detongnon J (2003) Cowpea yield gain from resistance to *Striga gesnerioides* parasitism in southern Benin. Exp Agric 39:327–333

Carvalho MHC, d'Arcy-Lameta A, Macauley HR, Gareil M, Maarouf HE, Pham-Thi AT, Zuily-Fodil Y (2001) Aspartic protease in leaves of common bean (*Phaseolus vulgaris* L.) and cowpea (*Vigna unguiculata* L. Walp): enzymatic activity, gene expression and relation to drought susceptibility. FEBS Lett 492:242–246

Chen CY, Pan L, Hu ZH, Peng H, Ding Y (2008) RAPD analysis of yardlong bean (*Vigna unguiculata* ssp. *sesquipedalis*) cultivars. J Jianghan Univ 36:77–83

Chen X, Laudeman TW, Rushton PJ, Spraggins TA, Timko MP (2007) CGKB: an annotation knowledge base for cowpea (*Vigna unguiculata* L.) methylation filtered genomic gene space sequences. BMC Bioinformatics 8:129

Close TJ, Luo M, Lonardi S, Alpert M, Duma D, Bozdag S, Farmer A, May GD, Jackson SA, Ma Y, Gu YQ, Wu Jm You FM, Diop NN, Wanamaker S, Fenton RD, Timko MP, Roberts PA and Ehlers JD (2011) Genome resources for cowpea genotype IT97K-499-35. Abstracts plant & animals genomes XIX conference, January 11–15, 2011, Town & Country Convention Center, San Diego, CA, P028

Coetzer N, Gazendam I, Oelofse D, Berge DK (2010) SSH screen and SSHdb, generic software for microarray based gene discovery: application to the stress response in cowpea. Plant Methods 6:10

Coulibaly S, Pasquet RS, Papa R, Gepts P (2002) AFLP analysis of the phonetic organization and genetic diversity of cowpea (*Vigna unguiculata*) reveals excessive gene flow between wild and domesticated types. Theor Appl Genet 104:358–366

Das S, Bhat PR, Sudhakar C, Ehlers JD, Wanamaker S, Roberts PA, Cui X, Close TJ (2008) Detection and validation of single feature polymorphisms in cowpea (*Vigna unguiculata* L. Walp) using a soybean genome array. BMC Genomics 9:107

Das S, Ehlers JD, Close TJ, Roberts PA (2010) Transcriptional profiling of root-knot nematode induced feeding sites in cowpea (*Vigna unguiculata* L. Walp.) using a soybean genome array. BMC Genomics 11:480

Diop NN, Kidri M, Repellin A, Gareil M, d'Arcy-Lameta A, Pham-Thi AT, Zuily-Fodil Y (2004) A multicystatin is induced by drought-stress in cowpea [*Vigna unguiculata* (L) Walp.] leaves. FEBS Lett 577:545–550

Ehlers JD, Matthews WC, Hall AE, Roberts PA (2000) Inheritance of a broad-based form of nematode resistance in cowpea. Crop Sci 40:611–618

Fang JG, Chao CT, Roberts PA, Ehlers JD (2007) Genetic diversity of cowpea (*Vigna unguiculata*) in four West African and USA breeding programmes as determined by AFLP analysis. Genet Resour Crop Evol 54:1197–1209

Fatokun CA (1997) Breeding cowpea for resistance to insect pests: attempted crosses between cowpea and V. vexillata. In: Singh BB, Mohan Raj DR, Dashiell K, Jackai LEN (eds) Advances in cowpea research. Copublication of International Institute of Tropical Agriculture (IITA)

and Japan International Center for Agricultural Sciences (JIRCAS), IITA, Ibadan, Nigeria, pp 52–61
Fatokun CA, Danesh D, Menancio-Hautea DI, Young ND (1993) A linkage map for cowpea [*Vigna unguiculata* (L) Walp] based on DNA markers (2n = 22). In: O'Brien JS (ed) Genetic maps 1992. A compilation of linkage and restriction maps of genetically studied organisms. Cold Spring Harbor Laboratory Press, Cold Spring Harbour, NY, pp 6256–6258
Fatokun CA, Menancio-Hautea DI, Danesh D, Young ND (1992) Evidence for orthologous seed weight genes in cowpea and Mungbean based on RFLP mapping. Genetics 132:841–846
Gupta SK, Gopalakrishna T (2010) Development of unigene-derived SSR markers in cowpea (*Vigna unguiculata*) and their transferability to other *Vigna* species. Genome 53:508–523
Hall AE, Singh BB, Ehlers JD (1997) Cowpea breeding. Plant Breed Rev 15:215–274
Ismail AM, Hall AE (1998) Positive and potential negative effects of heat-tolerance genes in cowpea. Crop Sci 38:381–390
Iuchi S, Kobayashi M, Yamaguchi-Shinozaki K, Shinozaki K (2000) A stress-inducible Gene for 9-cis-Epoxycarotenoid Dioxygenase involved in Abscisic Acid Biosynthesis under Water Stress in Drought-Tolerant Cowpea. Plant Physiol 123:553–562
Iuchi S, Yamaguchi-Shinozaki K, Urao T, Terao T, Shinozaki K (1996) Novel drought-inducible genes in the highly drought-tolerant cowpea: cloning of cDNAs and analysis of the expression of the corresponding genes. Plant Cell Physiol 37:1073–1082
Jones-Rhoades MW, Bartel DP, Bartel B (2006) MicroRNAs and their regulatory roles in plants. Annu Rev Plant Biol 57:19–53
Kolawole GO, Tian G, Singh BB (2002) Differential response of cowpea lines to P fertilizer. In: Fatokun CA, Tarawali SA, Singh BB, Kormawa PM, Tamo M (eds) Challenges and opportunities for enhancing sustainable cowpea production. Ibadan, Nigeria, IITA, pp 319–328
Kongjaimum A, Kaga A, Tomooka N, Somta P, Shimizu T, Shu Y (2012) An SSR-based linkage map of yardlong bean (*Vigna unguiculata* L. walp. Subsp. *unguiculata* Sesquipedalis Group) and QTL analysis of pod length. Genome 55:81–92
Lee JR, Back HJ, Yoon MS, Park SK, Cho YH, Kim CY (2009) Analysis of genetic diversity of cowpea landraces from Korea determined by Simple Sequence Repeats and establishment of a core collection. Korean J Breed Sci 41:369–376
Li CD, Fatokun CA, Ubi B, Singh BB, Scoles GJ (2001) Determining genetic similarities and relationships among cowpea breeding lines and cultivars by microsatellite markers. Crop Sci 41:189–197
Li GJ, Liu YH, Ehlers JD, Zhu ZJ, Wu XH, Wang BG (2007) Identification of an AFLP fragment linked to rust resistance in asparagus bean and its conversion to a SCAR marker. Hort Sci 42:1153–1156
Li J, Timko MP (2009) Gene-for-gene resistance in *Striga*-cowpea associations. Science 325:1094
Li J, Lis KE, Timko MP (2009) Molecular genetics of race specific resistance of cowpea to Striga gesnerioides. Pest Manag Sci 65:520–527. doi:10.1002/ps. 1722
Lu Y, Yang X (2010) Computational identification of novel micro RNAs and their targets in Vigna unguiculata. Comp Funct Genom 2010:Article ID 128297. doi:10.1155/2010/128297, 17 pages
Lucas MR, Diop NN, Wanamaker S, Ehlers JD, Roberts PA, Close TJ (2011) Cowpea–soybean synteny clarified through an improved genetic map. Plant Genome 4:218–225
Luo MC, Thomas C, You FM, Hsiao J, Ouyang S, Buell CR, Malandro M, McGuire PE, Anderson OD, Dvorak J (2003) High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82:378–389
Maarouf HE, Zuily-Fodil Y, Gareil M, d'Arcy-Lameta A, Pham-Thi AT (1999) Enzymatic activity and gene expression under water stress of phospholipase D in two cultivars of *Vigna unguiculata* L. Walp. differing in drought tolerance. Plant Mol Biol 39:1257–1265
Malviya N, Sarangi BK, Yadav MK, Yadav D (2012) Analysis of genetic diversity in coepea (*Vigna unguiculata*) cultivars with random amplified polymorphic DNA markers. Plant Syst Evol 298:523–526

Menéndez CM, Hall AE, Gepts P (1997) A genetic linkage map of cowpea (*Vigna unguiculata*) developed from a cross between two inbred, domesticated lines. Theor Appl Genet 95:1210–1217

Myers GO, Fatokun CA, Young ND (1996) RFLP mapping of an aphid resistance gene in cowpea (*Vigna unguiculata*). Euphytica 91:181–187

Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S, Pottorff M, Hearne S, Cisse N, Fatokun CA, Ehlers JD, Roberts PA, Close TJ (2009a) A consensus genetic map of cowpea [*Vigna unguiculata* (L) Walp.] and synteny based on EST-derived SNPs. Proc Natl Acad Sci U S A 106:18159–18164

Muchero W, Ehlers JD, Close TJ, Roberts PA (2009b) Mapping QTL for drought stress induced premature senescence and maturity in cowpea (*Vigna unguiculata* (L.) Walp.). Theor Appl Genet 118:849–863

Muchero W, Ehlers JD, Close TJ, Roberts PA (2010) QTL analysis for resistance to foliar damage caused by *Thrips tabaci* and *Frankliniella schultzei* feeding in cowpea [*Vigna unguiculata* (L.) Walp.]. Mol Breed 25:47–56

Muchero W, Ehlers JD, Close TJ, Roberts PA (2011) genic SNP markers and legume synteny reveal candidate genes underlying QTL for *Macrophomina phaseolina* resistance and maturity in cowpea [*Vigna unguiculata* (L.) Walp.]. BMC Genomics 12:8–22

Ng NQ, Singh BB (1997) Cowpea. In: Fuccillo D, Sears L, Stapleton P (eds) Biodiversity and trust. Cambridge University Press, Cambridge, pp 82–89

Nielsen SS, Brandt WE, Singh BB (1993) Genetic variability for nutritional composition and cooking time of improved cowpea lines. Crop Sci 33:469–472

Nkongolo KK (2003) Genetic characterization of Malawian cowpea (*Vigna unguiculata*) landraces: diversity and gene flow among accessions. Euphytica 129:219–228

Ogunkanmi LA, Ogundipe OT, Ng NQ, Fatokun CA (2008) Genetic diversity in wild relatives of cowpea (*Vigna unguiculata*) as revealed by simple sequence repeats markers. J Food Agric Environ 6:263–268

Ouédraogo JT, Gowda BS, Jean M, Close TJ, Ehlers JD, Hall AE, Gillaspie AG, Roberts PA, Ismail AM, Bruening G, Gepts P, Timko MP, Belzile FJ (2002) An improved genetic linkage map for cowpea (*Vigna unguiculata L.*) combining AFLP, RFLP, RAPD, biochemical markers and biological resistance traits. Genome 45:175–188

Paul S, Kundu A, Pal A (2011) Identification and validation of conserved microRNAs along with their differential expression in roots of *Vigna unguiculata* grown under salt stress. Plant Cell Tiss Org Cult 105:233–242

Sanginga N, Lyasse O, Singh BB (2000) Phosphorus use efficiency and nitrogen balance of cowpea breeding lines in a low P soil of the derived savanna zone in West Africa. Plant Soil 220:119–128

Sawadogo M, Ouedraogo JT, Gowda BS, Timko MP (2010) Genetic diversity of cowpea (*Vigna unguiculata*) cultivars in Burkina- Faso resistant to *Striga gesnerioides*. African J Biotech 9:8146–8153

Simoes-Araujo JL, Rodrigues RL, de A Gerhardt LB, Mondego JMC, Alves-Ferreira M, Rumjanek NG, Margis-Pinheiro M (2002) Identification of differentially expressed genes by cDNA-AFLP technique during heat stress in cowpea nodules. FEBS Lett 515:44–50

Singh BB (2001) Genetic variability for physical properties of cowpea seeds and their effect on cooking quality. African Crop Sci Proc 5:43–46

Singh BB (2005) Cowpea [*Vigna unguiculata* (L.)Walp]. In: Singh RJ, Jauhar PP (eds) Genetic resources, chromosome engineering and crop improvement, vol 1. CRC Press, Boca Raton, FL, pp 117–162

Singh BB, Emechebe AM (1990) Inheritance of Striga gesnerioides: resistance in cowpea genotype B 301. Crop Sci 30:879–881

Singh BB, Matsui T (2002) Breeding cowpea varieties for drought tolerance. In: Fatokun CA, Tarawali SA, Singh BB, Kormawa PM, Tamo M (eds) Challenges and opportunities for enhancing sustainable cowpea production. IITA, Ibadan, Nigeria, pp 287–300

Singh BB, Sharma B (1996) Restructuring cowpea for higher yield. Indian J Genet 56:389–405
Singh BB, Ajeigbe HA, Tarawali SA, Fernandez-Rivera S, Abubakar M (2003a) Improving the production and utilization of cowpea as food and fodder. Field Crop Res 84:169–177
Singh BB, Ehlers JD, Sharma B, Freire-Filho FR (2002) Recent progress in cowpea breeding. In: Fatokun CA, Tarawali SA, Singh BB, Kormawa PM, Tamo M (eds) Challenges and opportunities for enhancing sustainable cowpea production. IITA, Ibadan, Nigeria, pp 22–40
Singh BB, Hartmann P, Fatokun CA, Tamo M, Tarawali SA, Ortiz R (2003b) Recent progress in cowpea improvement. Chron Horticult 43:8–12
Solleti SK, Bakshi S, Purkayastha J, Panda SK, Sahoo L (2008a) Transgenic cowpea (*Vigna unguiculata*) seeds expressing a bean α-amylase inhibitor 1 confers resistance to storage pests, bruchid beetles. Plant Cell Rep 27:1841–1850
Solleti SK, Bakshi S, Sahoo L (2008b) Additional virulence genes in conjunction with efficient selection scheme, and compatible culture regime enhance recovery of stable transgenic plants in cowpea via *Agrobacterium tumefaciens*-mediated transformation. J Biotech 135:97–104
Sunkar R, Chinnusamy V, Zhu J, Zhu JK (2007) Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. Trends Plant Sci 12:301–309
Tantasawat P, Trongchuen J, Prajongjai T, Seehalak W, Jittayasothorn Y (2010) Variety identification and comparative analysis of genetic diversity in yardlongbean (*Vigna unguiculata* spp *sesquipdalis*) using morphological characters, SSR and ISSR analysis. Sci Hortic 124:204–216
Timko MP, Gowda BS, Ouedraogo J, Ousmane B (2007) Molecular markers for analysis of resistance to *Striga gesnerioides* in cowpea. In: Ejeta G, Gressell J (eds) Integrating new technologies for *Striga* control: towards ending the witch-hunt. Singapore, World Scientific Publishing Co. Pte Ltd, pp 115–128
Timko MP, Rushton PJ, Laudeman TW, Bokowiec MT, Chipumuro E, Cheung F, Town CD, Chen X (2008) Sequencing and analysis of the gene-rich space of cowpea. BMC Genomics 9:103
Van Boxtel J, Singh BB, Thottappilly G, Maule AJ (2000) Resistance of (*Vigna unguiculata* (L.)Walp.) breeding lines to black eye cowpea mosaic and cowpea aphid borne mosaic poty virus isolates under experimental conditions. J Plant Dis Prot 107:197–204
Varshney RK, Close TJ, Singh NK, Hoisington DA, Cook DR (2009) Orphan legume crops enter the genomics era. Curr Opin Plant Biol 12:202–210
Xu P, Wu X, Wang B, Liu Y, Ehlers JD, Close TJ, Roberts PA, Diop NN, Qin D, Hu T, Lu Z, Li G (2011) A SNP and SSR based genetic map of Asparagus Bean (*Vigna unguiculata* ssp. *sesquipedialis*) and comparison with the broader species. PLoS ONE 6(1):1–8
Xu P, Wu X, Wang B, Liu Y, Qin D, Ehlers JD, Close TJ, Hu T, Lu Z, Li G (2010) Development and polymorphism of *Vigna unguiculata* ssp. *unguiculata* microsatellite markers used for phylogenetic analysis in asparagus bean (*Vigna unguiculata* ssp. *sesquipedialis* (L.) Verdc.). Mol Breed 25:675–684
Young ND (1999) A continuously optimistic vision for marker assisted breeding. Mol Breed 5:505–510
Yunes AA, Andrade MT, De Sales MP, Morais RA, Fernandes KVS, Gomes VM, Xavier-Filho J, De Andrade MT (1998) Legume seed vicilins (7S storage proteins) interfere with the development of the cowpea weevil (*Callosobruchus maculates* (F.)). J Sci Food Agric 76(1):111–116
Zannou A, Kossou DK, Ahanchede ZJ, Agbicodo E, Struik PC (2008) Genetic variability of cultivated cowpea in Benin assessed by random amplified polymorphic DNA. African J Biotechnol 7:4407–4414

Chapter 8
Advances in Greengram and Blackgram Genomics

J. Souframanien and P. Dhanasekar

Abstract Greengram [*Vigna radiata* (L.) Wilczek] and blackgram [*V. mungo* (L.) Hepper] (both 2n = 2x = 22) are important legume crops in Asia, where it is a major source of dietary protein for its predominantly vegetarian population. Various genomic resources have been developed to accelerate the marker assisted selection in these crops. Different types of markers such as RFLPs, RAPDs, AFLPs, SSRs and ISSRs have been used in discerning genetic diversity and developing linkage maps in these crops. In greengram, eight genetic linkage maps have been developed so far but no map contained enough markers to resolve all the 11 linkage groups, while the two linkage maps constructed in blackgram resolved all 11 linkage groups. Markers have been used for tagging and mapping of genes and QTLs for resistance against mungbean yellow mosaic virus, powdery mildew and *Cercospora* leaf spot diseases, bruchids and for seed traits. Comparative genome mapping between greengram and several other legumes including azuki bean, common bean, cowpea, soybean and lablab revealed various levels of macrosynteny depending on species, with the greatest upon common bean. Comparison between blackgram and azuki bean maps revealed high degree of genome colinearity. Efforts have been made in developing BAC libraries in greengram to facilitate map based cloning of genes and QTLs. High throughput sequencing technologies have led to the partial nuclear genome sequencing (100 Mb) and complete sequencing of chloroplast and mitochondrial genomes of greengram. Annotation of transcriptome sequences for functional genes has been carried out in greengram. The ESTs and genomic data base from closely related legumes will be helpful in developing high throughput markers such as SSRs and SNPs. These resources have the potential to accelerate gene discovery, mapping and assist molecular breeding in these crops.

J. Souframanien, Ph.D. (✉) • P. Dhanasekar, M.Sc. (Ag.)
Nuclear Agriculture and Biotechnology Division, Bhabha Atomic Research Centre, Trombay, Mumbai, Maharashtra 400094, India
e-mail: souf@scientist.com

S. Gupta et al. (eds.), *Legumes in the Omic Era*, DOI 10.1007/978-1-4614-8370-0_8,

Keywords Genome • Linkage maps • QTLs • Molecular marker • Sequencing • Comparative mapping

Introduction

Greengram (*Vigna radiata* (L.) Wilczek) and blackgram (*V. mungo* (L.) Hepper) are important legume crops widely cultivated in Asia and in particular India, wherein they complement cereal-based diets with a large proportion of digestible protein through use as a pulse. The annual world area under greengram production is about 5 Million ha of which about 90 % is in Asia (Malik 1996). India is the biggest producer of greengram and blackgram where about 3.55 mha and 3.26 mha respectively were cultivated with a production of 1.8 mt and 1.74 mt respectively during 2010–2011 (Gupta 2012). While substantial yield improvements have been made in these crops, their yields are still low that has restricted their wider use as an alternative pulse crop in Asian farming systems. In this chapter, we provide an up to date review of genomic studies conducted on these two crops.

Genome Size

Greengram and blackgram, the diploid legumes with 2n = 2x = 22 are constituted under the subfamily Papilionoideae, clade Millettioid, genus *Vigna* Savi, subgenus *Ceratotropis*, section *Ceratotropis*. These crops like most of the other *Vigna* species have modest genome sizes estimated to be 0.60 pg/IC (579 Mbp) and 0.59 pg/IC (574 Mbp) respectively (Arumuganathan and Earle 1991). The species from different clades show genetic proximity owing to genome conservation depending on their phylogenetic relationships and such conserved or orthologus regions play a pivotal role in exploiting the genomic resources in under studied or orphan crops related by descent.

Genomic Resources

A broad range of genomic resources is available and can be used to accelerate legume improvement. These include marker reportier, expression sequence tag (EST) database, genome sequences (whole or partial), physical maps, molecular maps, DNA chips and bacterial artificial chromosome (BAC) or similar genomic libraries.

Molecular Markers

The lack of sequence information in greengram and blackgram has limited the advance of topical and robust molecular markers such as SSR, ESTs and SNPs in these crops. In this section, the availability of different molecular markers in greengram and blackgram have been comprehensively discussed.

In greengram, RFLP markers have been used to map (Young et al. 1993) or identify a new source of resistance (Chaitieng et al. 2002) to powdery mildew disease. Humphry et al. (2003) identified RFLP markers linked to major powdery mildew resistance locus while, Fatokun et al. (1992) studied orthologous seed weight genes using RFLP. Besides, requiring large quantity of DNA for analysis, the time and labor intensive RFLP requires radioactive labeled probes that limit their wide application in spite of their high polymorphism. RFLP markers of both cDNA and random genomic clones of greengram were reported by Young et al. (1992). These RFLPs together with those from common bean, cowpea and soybean have been extensively used in greengram and or blackgram genome mapping. Souframanien et al. (2003) studied intra and inter-specific variations in the rDNA internal transcribed spacer (ITS) region using RFLP and found no variation within cultivated *V. mungo* species while inter-specific variation was detected among wild *Vigna* species.

RAPD markers have been used for the identification of greengram cultivars and for assessing the genetic diversity (Santalla et al. 1998; Lakhanpaul et al. 2000). Assessment of genetic diversity using RAPD analysis shows close similarity among greengram cultivars (Lakhanpaul et al. 2000). The study revealed narrow genetic base of Indian cultivars probably due to repeated use of limited ancestors in their pedigrees. This observation has further been confirmed using RAPD (Afzal et al. 2004; Betal et al. 2004). Significant polymorphism among gamma ray induced mutants has been observed using RAPD (25.8 %) and ISSR (33.3 %) markers in blackgram (Souframanien et al. 2002).

Yu et al. (1999) reported the abundance and variation of microsatellite DNA sequences in *Phaseolus* and *Vigna*. The cross amplification of soybean SSRs in *Vigna* species was studied by Peakall et al. (1998) and they found that there was 3–13 % cross amplification. Only recently microsatellite or SSR markers have been developed from greengram (Kumar et al. 2002a, b; Miyagi et al. 2004; Gwag et al. 2006). However, SSRs from azuki bean, common bean and cowpea can be used in both greengram and blackgram. As high as 72.7 % and 78.2 % of the azuki bean SSRs amplify greengram and blackgram genomic DNA, respectively, while 60.6 % of common bean SSRs amplify greengram genomic DNA (Souframanien and Gopalakrishna 2009). Gupta and Gopalakrishna (2009) demonstrated that the azuki bean microsatellite markers are highly polymorphic and informative and could be successfully used for genome analysis in blackgram. Gupta and Gopalakrishna (2010) reported the transferability of functional unigene-derived SSR markers in cowpea (*Vigna unguiculata*) to other *Vigna* species including greengram and blackgram.

ISSR markers have been successfully utilized for analysis of repeat motifs in greengram (Singh et al. 2000), genetic relationships in the genus *Vigna* (Ajibade et al. 2000), varietal identification in blackgram (Ranade and Gopalakrishna 2001). Singh (2003) revealed narrow genetic base of Indian cultivars using ISSR. ISSR markers were comparatively more efficient than RAPD in assessing genetic diversity among blackgram cultivars (Souframanien and Gopalakrishna 2004). AFLP marker study also reiterated the low genetic diversity in greengram (Bhat et al. 2005). High polymorphism was obtained with +3 than with +2 primers. Saini et al. (2004) reported that long primers (18–22 bases) in comparison to the 10-mer primers could efficiently dissect the genetic diversity and relationships in greengram germplasm.

Sivaprakash et al. (2004) were of the view that AFLP markers were successful in discerning high level of genetic diversity among blackgram landraces. Gupta et al. (2008) utilized AFLP markers along with others in developing a genetic linkage map in blackgram. AFLP markers were also highly efficient in unraveling the genetic diversity in blackgram as is evident from the study of Gupta and Gopalakrishna (2009) wherein an average of 29.4 polymorphic bands per primer combination were obtained. In their study, each AFLP primer pair was able to distinguish on average 19 of the 20 genotypes and there were seven AFLP primer pairs that were able to discriminate all 20 genotypes.

In recent years, a large number of genes conferring disease resistance to a diverse spectrum of pathogens have been isolated from wide range of plant species. Most of these genes have conserved amino acid motifs. The most notable being the presence of nucleotide binding site (NBS) and leucine rich repeat (LRR). The candidate resistance genes have been used successfully to develop *RGA* markers. Basak et al. (2004) showed RGA linked to yellow mosaic virus tolerance in blackgram. RGA primers of cowpea have also been used to divulge the genetic diversity across different YMV resistant and susceptible genotypes of greengram and blackgram (Narasimhan et al. 2010).

Mapping Populations

The eight genetic linkage maps in greengram and two in blackgram published till date are based on populations derived from F_2 or RILs from inter-subspecific crosses (Table 8.1) and BC F_1 or RILs from inter-subspecific crosses (Chaitieng et al. 2006; Gupta et al. 2008) respectively. Of the three intersubspecific greengram crosses ('VC3890A' × 'TC1966', 'Berken' × 'ACC41' and 'TC1966' × 'Pagasa 7') involved in the development of linkage maps, 'VC3890A', 'Berken' and 'Pagasa 7' are cultivated greengram types from *V. radiata* ssp. *radiata* whereas; 'TC1966' and 'ACC41' are accessions of the wild progenitor *V. radiata* ssp. *sublobata*. In blackgram also the intersubspecific crosses involved a cultivated cultivar 'JP219132' (Chaitieng et al. 2006) or 'TU 94-2' (Gupta et al. 2008) of *V. mungo* var. *mungo* and a wild genotype of *V. mungo* var. *silvestris*. The size of the mapping populations involved in developing the linkage maps varied from 104 to 180 in blackgram as against 58 to 202 in case of greengram.

QTLs

Quantitative traits are under the control of specific regions of chromosomes referred to as the quantitative trait loci (QTLs). In greengram and blackgram QTLs associated with quantitative traits related to insect pest and disease resistance and seed related characters have been mapped with molecular markers.

Table 8.1 Linkage maps developed in greengram and blackgram

Year of publication	Pedigree	Marker type	Population size/generation	No. of loci mapped	Map length (cM)	No. of linkage groups	Average marker interval (cM)	Reference
Greengram								
1993	'VC3890A' (ssp. *radiata*) × 'TC1966' (ssp. *sublobata*)	RFLP	58 F2	172	1,570	14	9.0	Menancio-Hautea et al. (1993)
1995	'VC3890A' (ssp. *radiata*) × 'TC1966' (ssp. *sublobata*)	RFLP	58 F2	102	1,020	13	10.0	Boutin et al. (1995)
1995	'VC3890A' (ssp. *radiata*) × 'TC1966' (ssp. *sublobata*)	RFLP	58 F2	133	1,240	12	9.3	Boutin et al. (1995)
2000	'Berken' (ssp. *radiata*) × 'ACC41' (ssp. *sublobata*)	RFLP, RAPD	67 F2	110	758	12	6.8	Lambrides et al. (2000)
2000	'Berken' (ssp. *radiata*) × 'ACC41' (ssp. *sublobata*)	RAPD	67 RIL	115	692	12	6.0	Lambrides et al. (2000)
2002	'Berken' (ssp. *radiata*) × 'ACC41' (ssp. *sublobata*)	RFLP	80 RIL	255	738	13	3.0	Humphry et al. (2002)
2002	'TC1966' × 'Pasaga 7'	AFLP	76 RILs	103	655.5	9	10.7	Sholihin (2002)
2010	'Berken' and a highly bruchid-resistant wild type 'ACC41' (*V. radiata* subsp. *sublobata*)	SSR, RFLP, RAPD, STS	202 RIL	179	1,831.8	12	10.2	Dan et al. (2010)
Blackgram								
2006	JP219132 (cultivated *V. mungo* var *mungo*) × TC2210 (wild blackgram var. *silvestris*)	RFLP, SSR, AFLP and Morphological	180 BC F1	148	783	11	6.4	Chaitieng et al. (2006)
2008	TU 94-2 (var *mungo*) × wild species (var. *silvestris*)	AFLP, SSR, ISSR and RAPD	104 RIL	428	865.1	11	2	Gupta et al. (2008)

Table 8.2 DNA markers and their nucleotide sequence linked with MYMV resistance gene in blackgram and greengram

Primer code	Sequence 5′–3′	Marker size (bp)	References
RGA-1-F-CG	AGTTTATAATTCGATTGCT	445	Basak et al. (2004)
RGA1-R	ACTACGATTCAAGACGTCCT		
RGASF1	GGNAAGACGACACTCGCNTTA	456	Maiti et al. (2011)
RGASR1	GACGTCCTNGTAACNTTGATCA		
RGA22F2	GGGTGGNTTGGGTAAGACCAC	1,236	Maiti et al. (2011)
RGA24R2	NTCGCGGTGNGTGAAAAGNCT		
ISSR811	GAGAGAGAGAGAGAGAC	1,357	Souframanien and Gopalakrishna (2006)
YMV1F	GAGAGAGAGAGAGAGACAAAG	1,357	
YMV1R	GAGAGAGAGAGAGAGACAGGA		
OPS-07	TCCGATGCTG		Selvi et al. (2006)

(a) *Mungbean yellow mosaic virus resistance:* Mungbean Yellow Mosaic Virus (MYMV), a whitefly (*Bemisia tabaci*) transmitted gemini virus belonging to the begomovirus having monopartite (one ~2.9 kb DNA) or bipartite genome (two ~2.6 kb DNAs referred to as "DNA-A" and "DNA-B") causes disease in a number of leguminous crops in India and South East Asia especially greengram, blackgram and soybean. Basak et al. (2004) used six YMV tolerant blackgram lines ('VM1'-'VM6') developed from a highly susceptible genotype 'T-9' and F_2 population to identify DNA markers linked to YMV tolerance. A RGA primer pair RGA-1-F-CG/RGA1-R amplified a 445 bp fragment only in homozygous tolerant and the heterozygous lines differentiating the YMV tolerant and susceptible parents and was found to be linked to YMV tolerance. The 445 bp marker was sequenced and named 'VMYR1'. The predicted amino acid sequence showed highly significant homology with the NB-ARC domain present in several gene products involved in plant disease resistance, nematode cell death and human apoptotic signaling. On further evaluation of more RGA primer pairs Maiti et al. (2011) identified two markers amplified using RGA primer pairs (RGASF1/RGASR1 and RGA22F2/RGA24R2) referred as YR4 and CYR1 linked to MYMV resistance (Table 8.2). Both these resistance linked markers are part of the open reading frames (ORFs) and possess conserved motifs of the NB-ARC domain having sequence homology with other virus resistance genes. Both the markers were validated using greengram and blackgram genotypes by multiplex PCR and showed YR4 to be partially linked and CYR1 to be completely linked. Souframanien and Gopalakrishna (2006) identified a tightly linked ISSR marker ($ISSR811_{1357}$) using a RIL mapping population (F_8) that was 6.8 cM away from the MYMV resistance gene loci. Sequence characterized amplified region (SCAR) primers designed (YMV1-F and YMV1-R) from this ISSR marker distinguished the MYMV resistant and susceptible plants in RIL population, agreeing well with the phenotypic data. The $ISSR811_{1357}$ marker was also validated using diverse blackgram genotypes differing in their MYMV reaction. In greengram, F_2 population derived from a cross between 'ML267'

and 'CO-4' was screened with RAPD primers and identified one marker $OPS7_{900}$ to be associated with YMV resistance (Selvi et al. 2006).

(b) *Powdery mildew resistance:* One of the main foliar diseases that affect the production of greengram and blackgram is powdery mildew; caused by fungus *Erysiphe polygoni* D.C. Severe infection by powdery mildew occurs in cool, dry months when it can reduce the yield of greengram by 20–40 % (Reddy et al. 1994). Molecular marker studies of powdery mildew in greengram have indicated both qualitative and quantitative inheritance. Young et al. (1993) used RFLPs to map genes in greengram that confer partial resistance to the powdery mildew fungus. F_3 lines derived from a cross involving cross between a moderately powdery mildew resistant ('VC3980A') and a susceptible ('TC1966') greengram parent were assayed in the field for powdery mildew response and the results were compared to the RFLP genotype data, thereby identifying powdery mildew response associated loci. A total of three genomic regions were found to have an effect on powdery mildew response, together explaining 58 % of the total variation. One marker showing a strong association with powdery mildew response was sgK472, located on LG 3 of greengram. The study of Humphry et al. (2003) identified a major locus conferring powdery mildew resistance in line 'ATF 3640'. 147 F_7 and F_8 RILs derived from a cross between 'Berken' (highly susceptible) and 'ATF 3640' (highly resistant) were screened for powdery mildew under glasshouse condition. RFLP linkage map constructed with 52 loci generated by 51 probes were used to identify a single major locus flanked by markers LpCS82 and VrCS73 on linkage group K. This locus peaked approximately 1.3 cM from marker VrCS65, explained 86 % of the total variation in the resistance response to the pathogen. However, location of this QTL did not coincide with any QTLs reported by Young et al. (1993). Kasettranan et al. (2010) identified two QTLs controlling the disease resistance in a RIL population of 190 F_7 lines. The population was developed from the cross between a susceptible cultivar, 'Kamphaeng Saen 1' and a resistant line, 'VC6468-11-1A'. Reaction to the disease was evaluated for resistance in field and greenhouse conditions. Analysis of variance revealed that 15 SSR loci on three linkage groups were associated with the resistance. Composite interval mapping consistently identified two QTLs on two LGs, *qPMR-1* and *qPMR-2*, conferring the resistance. *qPMR-1* and *qPMR-2* accounted for 20.10 % and 57.81 % of the total variation for plant response to the disease, respectively. Comparison based on common markers used in previous studies suggested that *qPMR-2* is possibly the same as the major QTL reported earlier using another resistant source. The SSR markers closely linked to *qPMR-1* (CEDG282 and CEDG191) and *qPMR-2* (MB-SSR238 and CEDG166) are useful in MAS for greengram powdery mildew resistance (Table 8.3).

(c) *Cercospora leaf spot disease resistance:* One of the most important diseases affecting greengram production in Asia is the *Cercospora* leaf spot (CLS), a foliar disease caused by the biotrophic fungus *Cercospora canescens* Illis & Martin (Chupp 1953). The fungus initially causes spotting on greengram leaves; the spots increase in number and size during flowering, but the increment is

Table 8.3 PCR markers and their nucleotide sequence linked with powdery mildew resistance gene in greengram

Primer/probe code	Sequence 5′–3′	Reference
VrCS SSR1F	GCGAAGTGATCTTATCTGCT	Zhang et al. (2008)
VrCS SSR1R	GTCAAATCTGAACCATAAA	
VrCS SSR2F	GTTGAAAACTACAATACACT	
VrCS SSR2R	ACCAACAGTTCCATATCATG	
VrCS SSR3F	GCAGACACAACCATAAATCC	
VrCS SSR3R	GGTCTTTGACGGCAATCTC	
VrCS STS1F	ATTACTTGAGGTGGGGATAAT	
VrCS STS1R	AATAGACCACTTTTTCCGT	
VrCS STS 2F	ATTTGATGACGATGTATTTAA	
VrCS STS2R	TAAAGATAATGCTGAGGG	
CEDG282F	CAGCAACAAGACATGGAGTG	Kasettranan et al. (2010)
CEDG282R	GGTGACCACTTAGACAGAC	
CEDG166F	GGTACAACATTCTTCTATTTG	
CEDG166R	GGCTTATGAGTTTATCTTATC	
MB-SSR238F	AGCTATTGGTGCATAGGTTC	
MB-SSR238R	GATATGATGAGTATGGTGTAG	
CEDG191F	CAATAAGCAATCTGTGGAGAG	
CEDG191R	CTGCAGGAAACTTGGAATTGC	

most rapid at the pod-filling stage. In susceptible varieties, infection expands rapidly resulting in premature defoliation and reduction in size of pods and seeds , and thus cause yield losses of up to 50 % if devoid of protection (AVRDC 1984). The progress in selecting CLS-resistant genotypes in large breeding programs is still limited. This is mainly due to the fact that CLS occurs only in the rainy season, which is the primary growing season for greengram. Moreover, field evaluation for resistance can be done in only one season per year, albeit greengram is considered among the shortest season crop in the world and can be grown 3–4 times a year. Molecular markers linked to the gene controlling resistance can aid selection and advance the generation year-round (Collard and Mackill 2008*)*. Chankaew et al. (2011) identified QTL for *Cercospora* resistance using F_2 ('KPS1' × 'V4718') and BC_1F_1 [('KPS1' × 'V4718') × 'KPS1'] populations developed from crosses between the CLS-resistant greengram 'V4718' and CLS-susceptible cultivar 'Kamphaeng Saen 1' (KPS1). CLS resistance in F_2 and BC_1F_1 populations was evaluated under field conditions during the wet seasons. Sixty nine polymorphic SSR markers were analyzed in the F_2 and BC_1F_1 populations. Segregation analysis indicated that resistance to CLS is controlled by a single dominant gene. Single regression analysis in the F_2 and BC_1F_1 identified seven SSR markers, namely CEDC031, CEDG044, CEDG084, CEDG117, CEDG305, VR108 and VR393, associated with CLS resistance ($P < 0.01$). All of them were located on LG 3, except CEDG044 which was located on LG 11. The R^2 of the markers ranged from 6.11 % (CEDG044) to

80.81 % (CEDG117). While composite interval mapping consistently identified one major QTL (*qCLS*) for CLS resistance on LG 3 in both F_2 and BC_1F_1 populations, *qCLS* was located between markers CEDG117 and VR393 on LG 3 at 26.91 and 24.91 cM. It accounted for 65.5–80.53 % of the disease score variation depending on seasons and populations. An allele from 'V4718' increased the resistance. The SSR markers flanking *qCLS* will facilitate transfer of the CLS resistance allele from 'V4718' into elite greengram cultivars.

(d) *Bruchids resistance*: Bruchids or seed weevils belonging to the genus *Callosobruchus* are the most important storage pest of pulses. Seeds of the leguminous crops are severely affected by bruchid species. The most serious of these species in Asia are azuki bean weevil (*C. chinensis* L.), cowpea weevil (*C. maculatus* F.) and graham bean weevil (*C. analis* F.). These three bruchid pest have different distribution ranges. *C. chinensis* occurs in Asia, where it is a pest on azuki bean, chickpea, cowpea, greengram, peanut, soybean and other grain legumes (Applebaum et al. 1969). *C. analis* occurs in Africa and Asia. *C. maculatus* is the most widely distributed of the bruchid species, occurring in Africa, Asia, and Australia. The genes responsible for bruchid resistance in two wild greengram strains, 'TC1966' and 'ACC41' have been mapped. The *Br* gene conferring resistance to *C. chinensis* in TC1966 that was initially mapped on LG 8 flanked by RFLP markers sgA882 and mgM151 at a distance of 3.6 cM and 6.5 cM respectively, was subsequently mapped to LG 9 at a distance of 0.2 cM with the marker Bng143 and 0.9 cM with the marker Bng110. Two point linkage analysis of RFLP marker data from F_2 DNA showed significant association of six RFLPs (pA352, pR26, pA882, pM151a, pA315, pA257). Bruchid resistance was located to a single locus on LG 8 between marker pA882 and pA315 at a distance of 3.6 cM and 27 cM away from the former and latter respectively. QTL mapping of bruchid resistance identified the only region on LG 8 to be significantly associated with a LOD value of 15.3, attributing 87.5 % of the total phenotypic variation (Young et al. 1992). Menancio-Hautea et al. (1993) constructed a RFLP linkage map of greengram and located bruchid resistance gene to a 13 cM interval flanked by RFLP markers. Bruchid resistance from 'TC1966', incorporated into cultivated greengram 'Osaka-ryokuto' conferred simultaneous inhibitory activity against the bean bug, *Riptortus clavatus* Thunberg and was characterized by the presence of a group of novel cyclopeptide alkaloids, called vignatic acids. Kaga and Ishimoto (1998) constructed a linkage map for *Br* and the vignatic acid gene (*Va*) using RAPD and RFLP probes developed from linked RAPD markers. *Va* cosegregated with bruchid resistance and mapped to single locus at the same position as the cluster of markers and 0.2 cM away from *Br.* Their finding suggests that a dominant gene or a cluster of genes controls the production of vignatic acids analogs. However, the study also showed vignatic acids producing lines with susceptible reaction suggesting that vignatic acids are not the principal factors involved in conferring resistance (Kaga and Ishimoto 1998). Kaga and Ishimoto (1998) showed eight RAPD markers were significantly associated and three RAPD (BEXA08, BEXA99 and BEXC49) to be tightly linked to resistance gene

Table 8.4 DNA markers and their nucleotide sequence linked with bruchid resistance gene in greengram

Primer/probe code	Sequence 5′–3′	Reference
pA882	RFLP probe	Young et al. (1992)
pM151a	RFLP probe	
pA315	RFLP probe	
BEXAO8	TTCGGACGAATA	Kaga and Ishimoto (1998)
BEXA99	GCGGTCAGCACA	
BEXC49	AGGGTGCGTATA	
pBEXA08	RFLP probe	
pBEXA99	RFLP probe	
pBEXc49	RFLP probe	
pBEXB32a	RFLP probe	
pBEXD02a	RFLP probe	
Bng143	RFLP probe	
SSRbr1F	ATGGGTAGCGTGATGCTG	Miyagi et al. (2004)
SSRbr1R	TGTCAAAATGTGGTTGGCG	
STSbr1F	CAGAAAACAAATCACAAGGC	
STSbr1R	GTAAGCATTGAAAAAGGGTG	
STSbr2F	CCACCCTATTCAATGCTTAC	
STSbr2R	ACACTTCAATGGCGGACG	
STSbr3F	CAAAAGTCCAACGCTGTTCCCTG	
STSbr3R	CCATCTGTGTAGAATCTCTCGGTG	
STSbr4F	GGTAAGGGTAGGGGTTTCCATTAG	
STSbr4R	GAGACAAAAAGAGGACCAAAGCC	
STSbr5F	TCAGTCTTCCGTTTACG	
STSbr5R	TTGAGTGCTCAGGGGA	
OPC-06	GAACGGACTC	Lei et al. (2008)
OPW02	ACCCCGCCAA	Chen et al. (2007)
OPW02aF	CCAAAGGAGTCGAGTGAAACT	
OPW02aR	GTTGTTGGGAAGGAGATA	
OPU11	AGACCCAGAG	
OPV02	AGTCACTCCC	
UBC223	GATCCATTGC	

(Table 8.4). The RFLP probes for these RAPD markers were used in RFLP mapping of bruchid resistance gene and RFLP markers 0.7 cM apart on either side of the *Br* gene was identified. Six markers pBEXA08, pBEXA99, pBEXc49, pBEXB32a, pBEXD02a and Bng143, were closest to the bruchid resistance gene, approximately 0.2 cM away.

Miyagi et al. (2004) successfully used BAC clones in greengram for the development of two PCR-based markers closely linked with a major locus conditioning bruchid (*C. chinensis*) resistance. These PCR based markers were validated in Indian *sublobata* accession ('Sub2') and twelve other greengram cultivars. Of the two STS primer pair used in their study, STSbr1 amplified 225 bp fragment in all the homozygous resistance plants tested (Sarkar et al. 2011). This tightly linked marker

Table 8.5 QTL analysis for *C. maculatus* resistance in a RIL population derived from *V. mungo* (cv. Tu 94-2) and *V. mungo* var. *silvestris* by composite interval mapping

Trait[a]	QTL name	Linkage group	Position (cM)	LOD score	Flanking markers	PVE[b] (%)	a(H)1[c]
CMRAE1.1	*Cmrae*1.1	LG 3	6.6	3.3	EACT/MCTC-8, EACT/MCAT-5	10.8	9.43
CMRAE1.2	*Cmrae*1.2	LG 4	32.4	5.1	CEDG086, CEDG154	16.3	11.84
CMRDP1.1	*Cmrdp*1.1	LG 1	3.9	3.4	CEDG133, CEDG149	9.8	4.87
CMRDP1.2	*Cmrdp*1.2	LG 1	63.3	4.0	EACG/MCTA-15, EAGG/MCTA-1	12.1	−5.65
CMRDP1.3	*Cmrdp*1.3	LG 2	39.0	2.7	OPL14-1300, OPI20-600	9.8	5.56
CMRDP1.4	*Cmrdp*1.4	LG 2	45.5	4.9	OPR1-1380, EACA/MCAT-12	16.4	6.51
CMRDP1.5	*Cmrdp*1.5	LG 2	54.3	2.9	EAAG/MCTA-1, UBC827-1800	10.3	5.09
CMRDP1.6	*Cmrdp*1.6	LG 10	80.5	2.8	EACG/MCTA-14, EAAG/MCAA-4	8.4	−4.66

Reprinted from Souframanien J, Gupta SK and Gopalakrishna T. Identification of quantitative trait loci for bruchid (*Callosobruchus maculatus*) resistance in black gram [*Vigna mungo* (L.) Hepper]. Euphytica 2010; 176:349–356. With permission from Springer Science + Business Media

[a]CMRAE *C. maculatus* adult emergence, CMRDP *C. maculatus* developmental period

[b]Percentage of phenotypic variance explained by QTL

[c]Additive effect

may be useful in generating superior genotypes with 'Sub2' bruchid resistance locus by marker assisted selection. Chen et al. (2007) developed 200 RILs (F_{12}) involving bruchid resistance accession 'TC1966' and MYMV resistant variety 'NM92'. Ten RAPD markers (UBC66, UBC 168, UBC 223, UBC 313, UBC 353, OPM04, OPU11, OPV02, OPW02 and OPW13) were found associated with the bruchid resistance through BSA. Four (OPW02, UBC223, OPU11 and OPV02) of these that were closely linked (Table 8.4) were cloned and transformed into SCAR and cleaved amplified polymorphism (CAP) markers. Seven codominant CAPs developed from the identified RAPD markers showed tighter linkage with the *Br* gene than the original RAPD.

In blackgram, an inter-subspecific mapping population (RIL) was generated by crossing *V. mungo* var. *mungo* (cv. TU 94-2, bruchid susceptible) and *V. mungo* var. *silvestris* (bruchid resistant). About 37.8 % of bruchids completed their lifecycle on seeds of *V. mungo* var. *silvestris* compared with 100 % on the susceptible variety TU 94-2. The total developmental period of *C. maculatus* on *Vigna mungo* var. *silvestris* was considerably extended (88 days as compared with 34 days on TU 94-2). A genetic linkage map constructed using RILs in F_9 generation with 428 markers [86 RAPD, 47 SSR, 41 ISSR, 254 AFLP] was used for QTL detection using one hundred four individuals. Two QTLs, *Cmrae*1.1 and *Cmrae*1.2, were identified for percentage adult emergence, on linkage group (LG) 3 and 4, respectively (Table 8.5). For developmental period, six QTLs were identified, with two QTLs (*Cmrdp*1.1 and *Cmrdp*1.2) on LG 1, three QTLs (*Cmrdp*1.3, *Cmrdp*1.4, and *Cmrdp*1.5) on LG 2 and one QTL (*Cmrdp*1.6) on LG 10 (Souframanien et al. 2010).

(a) *Seed weight*: The preference for large seed types by consumers has resulted in a sixfold increase in the seed weight/seed size of modern greengram varieties compared to the wild progenitor ssp. *sublobata* (Fatokun et al. 1992; Lambrides and Imrie 2000). Large seed size has therefore become an important trait in greengram breeding programs. Independent QTLs (4–11) associated with seed weight have been identified in mapping studies (Fatokun et al. 1992; Humphry et al. 2005). Fatokun et al. (1992) identified seed weight QTLs on LGs 1, 2, 3 and 4 accounting for 49 % of the phenotypic variation using 58 F_3 families. Orthologous genes for seed weight with a QTL of large effect were also identified on homologous genomic regions of LG 2 in both greengram and cowpea, which was later shown to be associated with seed weight in pea, *Pisum sativum* L. (Timmerman-Vaughan et al. 1996). Humphry et al. (2005) mapped 11 QTLs conditioning seed weight using 227 RILs accounting 80 % of the phenotypic variation, both under field and glasshouse trials with seven loci being common in both the datasets. Of the seven common QTLs, one was located on each of LGs 1, 5, 9, 10, 11 and two QTLs were located on LG 2. None of the loci appeared to co-localise with any of the QTLs identified by Fatokun et al. (1992) although several QTLs did map to equivalent linkage groups that could be attributed to the use of different genetic material and/or small population size (58 individuals) in the study of Fatokun et al. (1992).

(b) *Hard-seededness*: Hard-seededness plays an important role in contributing to the development of weather-tolerant varieties (Imrie et al. 1991; Williams 1989). Four QTLs for hard-seededness from field data and a single QTL from glasshouse data were mapped using 227 RILs by Humphry et al. (2005). The single QTL from the glasshouse data (hsA) co-localised with one of the QTLs from the field data on LG 1 (=LG K) explaining up to 23 % of the variation in the field trial and 11 % of the variation in the glasshouse trial. Lambrides (1996) also mapped a hard-seededness locus near hsA on LG 1 in an F_2 population of 'Berken' × 'ACC41'. These results support the previous observations that very few genes appear to control hard-seededness in greengram (Lambrides 1996; Lawn et al. 1988; Williams 1989). The inheritance pattern of hard-seededness in this population is similar to that observed in soybean, where one major QTL was identified which explained 30 % of the variation with other minor QTLs contributing to the quantitative distribution of phenotypes (Keim et al. 1990).

(c) *Seed appearance traits*: Seed quality is determined by several traits other than seed weight and hard-seededness in greengram. Appearance of the seed is determined by the presence or absence of the texture layer, the pigmentation of the texture layer and the color of the testa. These traits have been mapped in the 'Berken' × 'ACC41' population using 67 F_2 and 67 RIL individuals (Lambrides 1996; Lambrides et al. 2000). *A* continuous range of phenotypes was observed when individuals of the population 'Berken' (shiny seed) × 'ACC41' (dull seed) were scored for the amount of texture layer using a scoring system of 0 (shiny seed, no texture layer) to 5 (deep texture layer) (Lambrides 1996). QTL analysis detected three regions on LGs 1-3, 2 and 8-9, collectively accounting for 71.5 % of the phenotypic variation for the texture layer score. The QTL near pO9b

(LGs 8 and 9) was detected at LOD = 5.94, where the allele from 'ACC41' showed dominant gene action for deep texture layer. Unexpectedly, the gene action at a QTL detected near marker pM78 (LGs 1-3) was suggestive of over-dominance that could explain the unexpected appearance of dull seeded segregants from hybrids between shiny seed coat parents.

The green speckled black testa color of 'ACC41' dominant over green testa color was controlled by a single locus that mapped to LG 2 (Lambrides et al. 2000). RFLP marker pA235 was linked to the testa color locus at 2.4 cM (LOD = 15.0), and this region of the greengram genome was shown to segregate with distortion (Lambrides et al. 2004). LG 2 also contained a single locus controlling the pigmentation of the texture layer in 'ACC41' that was dominant over no pigmentation and was unlinked to the testa color locus. RFLP marker pA204 was linked to the pigmentation locus at 6.4 cM (LOD = 8.0) (Lambrides et al. 2004).

Linkage Maps

The availability of molecular maps facilitates marker-assisted selection, mapbased cloning, and mapping of QTLs of agronomic importance in many crop plants. Genetic linkage maps have been constructed in blackgram (Chaitieng et al. 2006; Gupta et al. 2008) and greengram (Humphry et al. 2002; Lambrides et al. 2000). Eight molecular linkage maps for greengram have been published (Table 8.1). These maps were constructed from the data of F_2 or RIL populations from inter-subspecific crosses of 'VC3980' (cultivated) × 'TC1966' (wild from Madagascar) or 'Berken' (cultivated) × 'ACC41' (wild from Australia) using mainly RFLP and/or RAPD markers or 'TC1966' × 'Pagasa 7' (cultivated) using AFLP makers. The population size ranged from 58 to 202 plants. The maps differ in length (655.5–1,831.8 cM), number of markers (102–255 markers), number of linkage groups (LG) (09–14), and level (12–30.8 %) and regions of marker distortion. The most comprehensive map consists of 255 loci with an average 3 cM distance between the adjacent markers. However, none of the maps resolved to 11 LGs, which is the haploid chromosome number of greengram, for which many more markers are required to saturate the map and also the genome coverage of the markers need to be determined.

Genome research in blackgram has received far less attention than mungbean. Only two genetic linkage maps have been developed in this crop. But unlike greengram, both the maps resolved to 11 linkage groups owing to higher degree of saturation of markers. Chaitieng et al. (2006) reported the first genetic linkage map of blackgram with 148 marker loci assigned to the 11 linkage groups, which correspond to the haploid chromosome number. Subsequently, another linkage map was constructed with 428 molecular markers which spanned a total distance of 865.1 cM with an average marker density of 2 cM (Gupta et al. 2008).

Comparative Genomics

Comparative linkage maps facilitate mapping of orthologous sequences among closely related plant species or genera and aid in better understanding of the organization and evolution of plant genomes. The use of heterologous RFLP probes has facilitated a number of comparative genome studies with various species of the genus *Vigna* (Boutin et al. 1995; Humphry et al. 2002; Kaga et al. 1996, 2000; Menancio-Hautea et al. 1993). In some instances, there is greater homology between greengram and species in different genera compared to greengram and related species from the subgenus *Ceratotropis*. Chromosomal rearrangements could explain many of the differences between the greengram and cowpea genomes (Menancio-Hautea et al. 1993) and the greengram and azuki bean genomes (Kaga et al. 2000). These studies show that conserved blocks of genes appear on several LGs, although no entire LG was conserved between greengram and cowpea and greengram and azuki bean. The genetic complement of greengram and cowpea was similar at the nucleotide level, although copy number changed and the linear arrangement of conserved linkage blocks also changed (Menancio-Hautea et al. 1993). Interestingly, this study showed that there were greater differences between the greengram and cowpea genomes compared to sorghum and maize, which are in different genera. Prior to the taxonomic reorganisation of *Phaseolus* and *Vigna,* greengram had been placed in the genus *Phaseolus*. Not surprisingly, with respect to marker order and conserved LGs, greengram shows greater homology to common bean than to cowpea (Boutin et al. 1995) and azuki bean (Kaga et al. 1996, 2000). All greengram LGs consisted of one, two or three LGs of common bean with the average conserved linkage block of 36. 2 cM and the longest conserved linkage block of 103.5 cM occurring on LG 8 of greengram and LG K of common bean. A balanced translocation between greengram (LGs 2 and 6) and common bean (LGs F and H) was also detected. Greengram LGs 8, 9, 10 (Menancio-Hautea et al. 1993) were all composed of markers from L, G, K of common bean, suggesting that these greengram LGs may represent segments of one LG. The studies of Kaga et al. (2000) and Lambrides et al. (2000) also provided evidence that greengram LGs 8 and 9 are segments of one LG. Although Menancio-Hautea et al. (1993) found that 88 % (125/142) of soybean genomic probes hybridized to greengram, there was substantial genome rearrangement between these two genomes (Boutin et al. 1995). Only short and scattered linkage blocks were conserved between greengram and soybean. For example, markers from up to 16 different LGs of soybean were found on LG 1 of greengram. The average conserved linkage block between the greengram and soybean genomes was about a third the size of conserved linkage blocks between greengram and common bean. There is surprising homology between greengram and a more distantly related member of *Fabaceae* lab lab (*Lablab purpureus*) (Humphry et al. 2002). Large conserved linkage blocks were observed between the two genomes, and linkage order was retained in the majority of cases, although evidence was presented to suggest the genomes to differ by at least one inversion and other complicated chromosomal rearrangements. Different copy numbers detected in each of the genomes suggested that they had also accumulated a large number of deletions/duplications after they diverged.

Gupta et al. (2008) compared the blackgram linkage map developed by them with that of the azuki bean map (Han et al. 2005) based on common SSR markers. The SSR markers were present on all 11 linkage groups with number per linkage group varying from 2 to 9. High level of colinearity was observed between the two maps. Forty-one SSR makers were shared between the two maps, the orders of which were highly conserved excepting five marker pairs. The markers were tightly linked but in opposite direction, suggesting few internal inversions that could have occurred during the evolutionary divergence of blackgram and azuki bean.

Most of the markers utilized in the development of blackgarm genome maps, especially SSRs and RFLPs, were previously mapped on azuki bean (Lee et al. 2001). Comparison of 80 common marker loci between the two maps revealed high degree (88 %) of genome colinearity. However, inversions, insertions, deletions, duplications and a translocation were also detected. For example, marker order on parts of LG 1, 2 and 5 is reversed between the two species. Significant macro- and microsynteny were observed among *G. max*, *P. vulgaris* and *Vigna radiata* (Lee et al. 2001). Large-scale macrosyntenic blocks were also observed among *P. vulgaris*, *M. truncatula*, and *L. japonicus* (McConnell et al. 2010). Because extensive genomic information is available for soybean (http://soybase.org/), *Medicago* (http://gbrowse.jcvi.org/cgi-bin/gbrowse/medicago/#search), and *Lotus* (http://www.plantgdb.org/LjGDB/), the genetic synteny between pulse and the model legume species will help pulse researchers to speed up the understanding of pulse genomes by comparative genomics (http://www.comparative-legumes.org/).

Genome Sequencing

Complete nuclear genome sequence information in greengram and blackgram is not available as on date and the use of genomic resources in these crops largely depend on the sequence information available in the closely related taxa. Progress towards the complete genome sequencing in these crops are at various stages and as a prelude Tangphatsornruang et al. (2009) has sequenced about 100 Mb of the greengram genome following shotgun sequencing. However, complete sequences of chloroplast (Tangphatsornruang et al. 2010) and mitochondrial genomes (Alverson et al. 2011) in greengram have been published. Among these two crops, more efforts have gone into greengram in comparison to blackgram as is evident from the number of ESTs established and the number of linkage maps available. With the advent of next generation sequencing, the task of whole genome sequencing of these crops can be efficiently completed.

Whole Genome Sequencing of Greengram

Tangphatsornruang et al. (2009) have generated and characterized a total of 470,024 genome shotgun sequences covering 100.5 Mb of the greengram genome using 454 sequencing technology. A total of 470,024 quality filtered sequence reads

was generated with the average read length of 216 bases covering 100.5 Mb. All reads can be retrieved from NCBI Short Read Archive (ID = SRA003681). The contig length ranges from 89 bases to 44,462 bases. The average GC content of greengram genomic DNA generated in this study is 34.69 % which is consistent with the reports on GC contents in other plant genomes such as *Arapbidopsis* (36 %). From the shotgun sequencing data, a total of 1,493 microsatellite regions were isolated. There were 889 dinucleotide repeats (DNPs), 282 trinucleotide repeats (TNPs), 123 tetranucleotide repeats (TTNPs), 124 pentanucleotide repeats (PNPs) and 75 SSRs with hexanucleotide repeats or more. The distribution of the number of motif repeat ranged from 4 to 30 repeats. The most common motif type of DNPs was TA/AT (89.3 % of DNPs) followed by TC/AG (7.1 % of DNPs) and AC/TG (3.6 % of DNPs). The GC/CG motif was not found in the data set. TNPs were found at 282 SSR loci (18.9 %), which was three times lower than that of DNPs. The TAA repeat was the most common motif type found at 184 loci (65.24 % of TNPs). The least frequent TNP motif was GC-rich (GCG/CGC) found at only two loci. The frequency of identified SSR in greengram was one SSR in every 67 kb (1,493 SSRs in 100.5 Mb) which is significantly lower than the SSR frequency in soybean (1/7.4 kb) (Cardle et al. 2000). Among plant species, the SSR frequencies range from 1/1.5 kb in coffee to 1/20 kb in cotton (Aggarwal et al. 2007; Cardle et al. 2000). The observed low SSR frequency was probably due to a large proportion of reads from the low coverage sequencing (0.2×) of the greengram genome were biased toward highly repetitive parts of the genome.

For sequence annotation and gene ontology, the contigs were analyzed to predict 44,112 Open Reading Frame (ORF) using *Medicago trunculata* as a model organism and default parameter conditions. For functional annotation, the potential coding regions were analyzed by BLAST2GO (Conesa et al. 2005) leading to consistent gene annotations, assigning gene names, gene products, EC numbers and Gene Ontology (GO) numbers. Gene Ontology provides a system to categorize description of gene products according to three ontologies: molecular function, biological process and cellular component. Sequence homology search revealed that there were 1,542 ORFs matches with non-redundant protein database with an E-value cut-off at E-6. Nine hundred and fifty sequences were mapped to one or more ontologies with multiple assignments possible for a given protein within a single ontology. There were 647 assignments made to the molecular function ontology, with a large proportion of these in catalytic (42.72 %) and binding activities (44.17 %) categories (Fig. 8.1a). Under the biological process ontology, 555 assignments were made with a large proportion of assignments falling into metabolic and cellular process (such as secretory pathway, transcription and translation) categories (Fig. 8.1b). Studies on similarity of greengram predicted ORFs with other plant ESTs showed that the greengram dataset and the *Glycine max* gene index gave the highest number of matched sequences (7,940 sequences). *V. radiata* and *G. max* are grouped together exhibiting extensive genome conservation based on previous comparative genetic mapping (Boutin et al. 1995; Choi et al. 2004). *M. truncatula*, which is a cold season legume, also shares a large number of homologous sequences (5,759 sequences) with the greengram dataset.

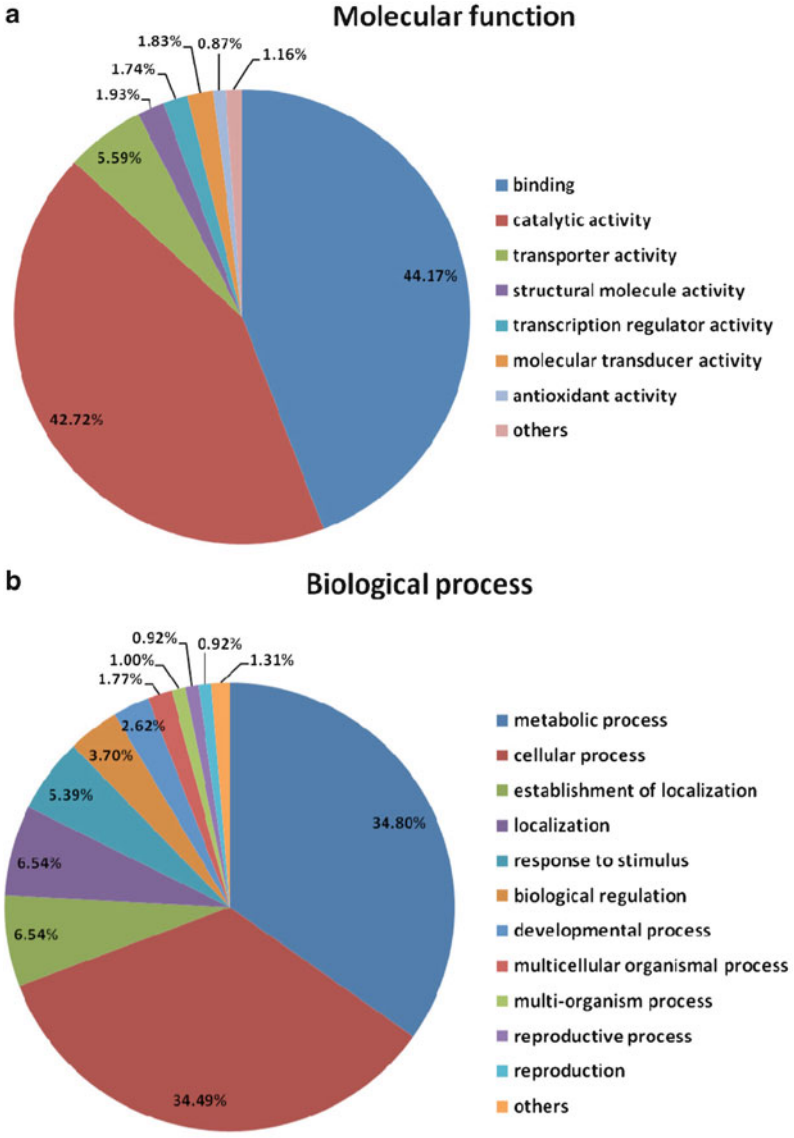


Fig. 8.1 Gene Ontology classification of the predicted mungbean ORFs according to molecular function (**a**) and biological process (**b**) using BLAST2GO with E-6 cutoff [Reprinted from Tangphatsornruang S, Somta P, Uthaipaisanwong P, Chanprasert J, Sangsrakru D, Seehalak W, Sommanas W, Tragoonrung S and Srinives P. Characterization of microsatellites and gene contents from genome shotgun sequences of mungbean (*Vigna radiata* (L.) Wilczek). BMC Plant Biol 2009; 9:137. With permission from BioMed Central, Inc.]

Greengram Chloroplast (cp) Genome

The complete cp genome sequence of greengram was described by Tangphatsornruang et al. (2010). This cp genome is 151,271 bp in length which includes a pair of inverted repeats (IRs) of 26,474 bp separated by a small single-copy region of 17,427 bp and a large single-copy region of 80,896 bp (Fig. 8.2). The genome contains 108 unique genes and 19 of these genes are duplicated in the IR. Of these, 75

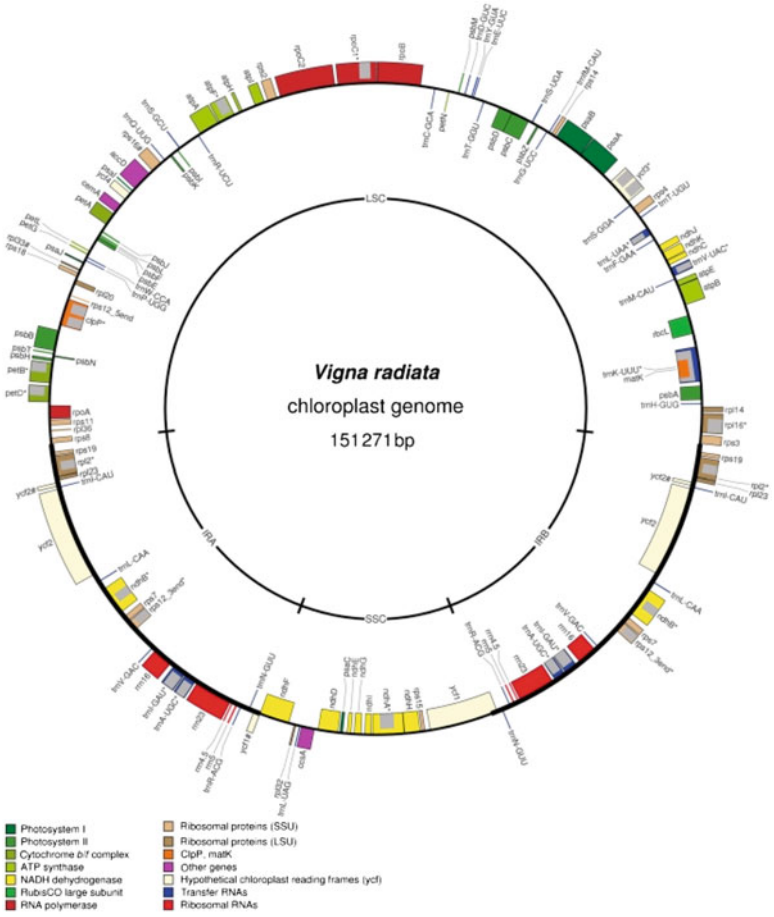


Fig. 8.2 Map of the *Vigna radiata* chloroplast genome [Reprinted from Tangphatsornruang S, Sangsrakru D, Chanprasert J, Uthaipaisanwong P, Yoocha T, Jomchai N and Tragoonrung S. The Chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. DNA Res 2010; 17:11–22. With permission from Oxford University Press]

are predicted protein-coding genes, 4 ribosomal RNA genes and 29 tRNA genes. The complete cp genome sequence was reported in the DDBJ/EMBL/GenBank nucleotide sequence database (GQ893027). Analysis of the repeat sequences in the greengram cp genome identified 22 direct repeats and 28 IRs of 30 bp or longer with a sequence identity of 90 %. Thirty repeats are 30–40 bp long, 11 repeats are 41–50 bp long, 4 repeats are 51–80 bp long and 5 repeats are longer than 80 bp. The longest direct repeat in greengram cp DNA is a 287-bp duplication of an internal fragment of *ycf2* ($_{\Psi}$ycf2) in the IRs which shared a very high sequence homology with those of *G. max* and *P. vulgaris*. Most of the direct repeats are distributed within the intergenic spacer regions, the intron sequences, and in the *trnS*, and *ycf2* genes. Sequence length polymorphism of 16 homopolymers among *V. radiata*, *V. unguiculata* (VU210 and TVNU294), *V. mungo* and *V. umbellata* were tested and no polymorphism was detected between varieties of *V. radiata*, although, the study observed polymorphism at the intra-specific level in *V. unguiculata*. This demonstrated that cp microsatellites reported could provide an assay for detecting polymorphism at the population-level and for comparison of more distant phylogenetic relationships at the genus level or above. These cp microsatellites can also be useful in ecological and evolutionary studies because they are non-recombinant, haploid and uniparentally inherited. The information from cp genomes be useful for studies of phylogenetic relationships, and will also facilitate cp transformation in greengram.

Greengram Mitochondrial Genome

The annotated mitochondrial genome sequence of greengram is available from GenBank (accession HM367685) (Alverson et al. 2011). This genome assembled into a single, circular-mapping molecule of length 401,262 nt and 45.1 % GC content, both of which are near the median values of fully sequenced seed plant mitochondrial genomes. The genome contains 31 protein, 3 rRNA, and 16 tRNA genes (Fig. 8.3). Two identical copies of the *atp9* gene are present in the genome. This is the most protein-gene-poor mitochondrial genomes so far sequenced in plants. This first completely sequenced legume mitochondrial genome also confirms the absence of the *cox2* gene supporting the best-studied case of recent functional transfer of an organellar gene to the nuclear genome, with the transfer restricted to a subset of papilionoid legumes. Although most other respiratory genes have never been found to have been lost during angiosperm evolution, 17 genes (15 ribosomal protein and 2 respiratory) are known to have been lost frequently. Nine of these 17 genes are either absent from the *Vigna* mitochondrial genome (*rpl2*, *rpl10*, *rps2*, *rps11*, *rps13*, *sdh3*) or are present as pseudogenes in various stages of attrition (*rps7*, *rps19*, *sdh4*). The *sdh4* gene is the most intact of these, with just a single 10-nt insertion located roughly 30 amino acids upstream of the conserved stop codon. The *Vigna*

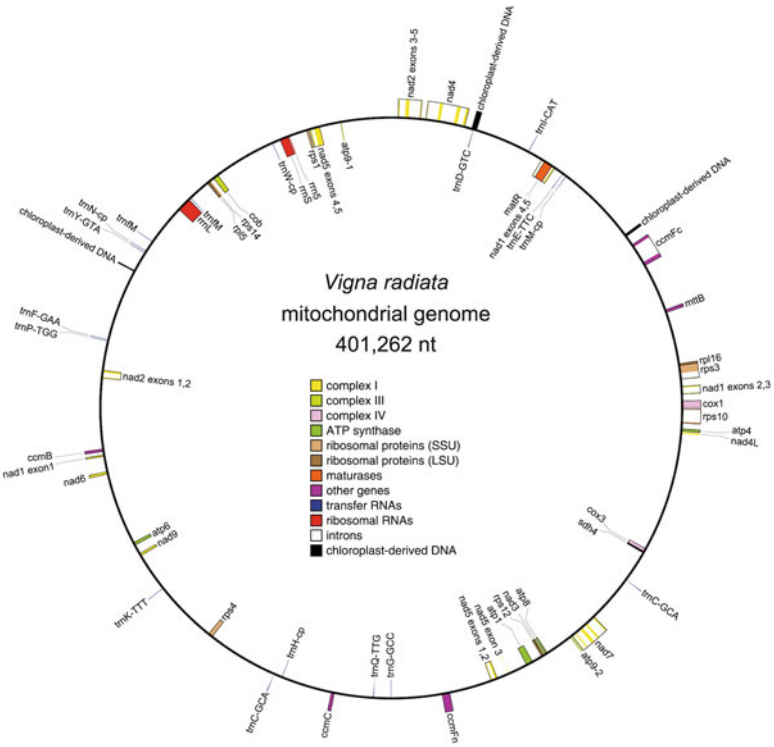


Fig. 8.3 Map of the *Vigna radiata* mitochondrial genome [Reprinted from Alverson AJ, Zhuo S, Rice DW, Sloan DB and Palmer JD (2011) The Mitochondrial Genome of the Legume *Vigna radiata* and the Analysis of Recombination across Short Mitochondrial Repeats. PLoS One 2011; 6:e16404. doi:10.1371/journal.pone.0016404. With permission from PLoS One]

mitochondrial genome contains a conserved set of 17 *cis*-spliced and five *trans*-spliced group II introns (Fig. 8.3). Seed plant mitochondrial genomes typically require *trans*-splicing of the intron separating exons 3 and 4 of the *nad5* gene to create a full-length *nad5* transcript. In *Vigna*, exon 3 is identically oriented and less than 3 kb apart from exon 4 (Fig. 8.3), raising the possibility of a recent reversion to *cis-splicing* of this intron. When the greengram mitochondrial sequence was analyzed for the repetitive DNA, fewer repeats contributed to overall size of the greengram genome (just 2.7 % coverage compared to 8–62 % coverage in other genomes). Most greengram mitochondrial repeats are less than 100 nt in length, and most of these are less than 40 nt in length. The largest repeat in the greengram mitochondrial genome contains a duplicate copy of the *atp9* gene, and at just 297 nt in length, is substantially shorter than the largest repeat in all other fully sequenced seed plant mitochondrial genomes.

Data Mining

EST-SSRs

The EST databases available for many crop species provide a valuable resource for the identification and development of SSR markers. The sequences available in these databases can be mined for SSR repeats, thereby reducing the time and cost in developing microsatellite-enriched libraries. EST-SSRs, being part of the genes, are more useful as genetic markers because they represent variation in the expressed portion of the genome. EST-SSRs have been developed in a large number of plant species including chickpea (Choudhary et al. 2009), soybean (Hisano et al. 2007), common bean (Hanai et al. 2007) and *Medicago* spp. (Eujayl et al. 2004). Like genomic SSR markers, EST-SSR markers could be used for a variety of applications such as molecular mapping, gene tagging, and genetic diversity analysis (Varshney et al. 2005). In addition, EST-SSR markers show a high rate of transferability to related species or genera owing to the higher conservation of expressed sequences across species (Varshney et al. 2005). Hence, SSR markers developed in one species can be used in related species for which sufficient sequence information is not available for marker development. However, owing to large redundancy in the public EST databases, multiple sets of markers can be developed for the same locus. This problem can be circumvented by clustering the ESTs into a non-redundant set of gene-oriented clusters called unigenes. Primer pairs successfully developed from cowpea unigene SSRs were demonstrated to show cross species amplification and polymorphism in greengram, blackgram and other *Vigna* species (Gupta and Gopalakrishna 2010). The unigene SSR markers developed in this study showed a high rate of transferability (88 %) to other *Vigna* species, indicating the conservation of microsatellite sequences in the genus *Vigna* during evolution. These SSR markers would be helpful in the development of a saturated genetic linkage map and tagging genes in greengram and blackgram.

Intron Length Polymorphism (ILP)

Like other molecular markers, ILP markers can be used for a variety of applications like molecular mapping, gene tagging, genetic diversity analysis and comparative studies. In addition, ILP markers show a high rate of transferability to related species owing to a higher conservation of EST sequences across species. ILP markers were developed from cowpea EST. One hundred and ten PCR primers targeting one or more introns were developed from randomly chosen cowpea EST sequences and showed cross species amplification and polymorphism in greengram, blackgram and other related *Vigna* species (Gupta et al. 2012). Based on the sequence information from cross species amplified ILP marker, it was also observed that the exonic regions

were highly conserved among the *Vigna* species and large differences, including length variation and point mutations, were observed in the intronic regions. Therefore, ILP markers developed in one species can be used in a related species for which sufficient genomic resources for marker development are not available.

Bacterial Artificial Chromosome (BAC)

BAC libraries have been widely used in different aspects of genome research. However, there are only few reports of BAC library in *Vigna* species. Miyagi et al. (2004) constructed two greengram BAC libraries that together gave a 3.5 × coverage of the 587 Mb genome. The libraries were constructed from both *radiata* ssp. (green gram) using genotypes 'ACC41' and 'ATT3640' and its wild progenitor *sublobata* ssp. (golden gram) by cloning the DNA in pBeloBacII vector with an average insert size of 107 and 113 kb size. Two PCR-based markers were developed closely linked to a major locus conditioning bruchid resistance, by screening these libraries using RFLP probes, including Mgm213 that is very closely linked (1.3 cM). This information should aid in the introgression of this resistance locus into agriculturally elite cultivars. These libraries could also facilitate development of other PCR-based markers linked to other desirable traits. In near future, the BACs of pulse crops should have potential applications in pulse comparative genomics and functional genomics as well owing to the macro- and microsynteny widespread within legumes.

Transcriptome Analysis of Greengram

Transcriptome analysis provided a powerful tool for differential gene expression, mutant splicing, SSR or SNP analysis, and functional genetics studies. The discovery of SSR and single nucleotide polymorphism (SNPs) markers based on transcribed regions has become a common application in plants because of the larger number of ESTs available in database. Moe et al. (2011) identified SSR and SNPs from greengram transcriptome sequencing. Two greengram genotypes ('Sunhwa' and 'Jangan') transcriptome sequencing yielded 61.71 Mb and 60.68 Mb with 411 and 424 per read by 150,159 and 142,993 reads for 'Sunhwa' and 'Jangan', respectively. A total of 5,254 and 6,374 large contigs (≥500 bp), with an average length of 833 and 853 were detected for 'Sunhwa' and 'Jangan', respectively. Approximately 41 % (8,606) of 'Sunhwa' ESTs and 41.74 % (10,758) of 'Jangan' ESTs were matched to known functional sequences. A cluster analysis revealed correlations among the transcriptome profiles. Functional categories such as structural or catalytic proteins with binding function or cofactor requirements, subcellular localization, metabolism, protein fate, regulation of metabolism and protein function, and cellular transport were dominantly represented in the young-leaf greengram transcriptome, whereas genes corresponding to energy, cell cycle and DNA processing, transcription,

protein synthesis, cellular communication/signal transduction mechanisms, cell rescue, defense and virulence, environmental interaction, systemic interaction with the environment, cell fate, systemic development, cellular component biogenesis, and organ differentiation represented a lower percentage.

On analyzing the greengram transcriptome for SSR and SNPs, it was reported that tri-nucleotide 743 (55.7 %) and 915 (56.1 %) type SSR motifs were most abundant, followed by di-nucleotides 448 (33.9 %), 552 (33.9 %), and others 143 (10.7 %), and 163 (10.0 %) in both 'Sunhwa' and 'Jangan'. Of the tri-nucleotide types, the GAA/AAG/AGA class dominated (226 and 284), whereas the GA/AG class dominated (255 and 377) in the di-nucleotide types in both 'Sunhwa' and 'Jangan'. The different features of the repeat motif types present in the transcriptome sequences indicated that some differences exist between these two greengram varieties. On the other hand, greengram genomic DNA sequence was reported to have more of dinucleotide (TA/AT) repeats (Tangphatsornruang et al. 2009). Other than the differences in the SSR repeat motifs, variations in the SNPs in these two varieties were also observed. Assembly using GS Reference Mapper software revealed 8,249 SNP variations, which was supported by the 69,915 read count. Among all variations, the maximum value was an indel (1–94 nucleotide) (22.2 %), with 141 (>6 nucleotide indel) included in it. It was then followed by C/T (11.53 %), T/C (11.25 %), G/A (10.49 %), A/G (10.45 %), others (5.9 %), A/T (4.49 %), T/A (4.16 %), A/C (3.36 %), C/A (3.33 %), G/C (3.33 %), G/T (3.31 %), C/G (3.12 %), T/G (2.86), and N/K (0.23 %). Next-generation transcriptome sequencing will serve as a superior resource for developing polymorphic DNA markers, not only because of the enormous quantities of sequence data in which markers can be discovered, but also because the discovered markers are gene-based. Such markers are advantageous because they facilitate the detection of functional variation and selection in genomic scans or genetic association studies.

EcoTILLING of Greengram

EcoTILLING has been demonstrated as a powerful tool to uncover SNPs and their approximate location without having to sequence all individuals in the population. This technique is especially useful when working with plants that have a narrow genetic base and looking for variation in highly conserved genes in mungbean (Barkley et al. 2008). DNA from a single reference plant *V. radiata* var. *sublobata* was mixed with that of each member of the greengram population to mine for polymorphic sites. Amplifying DNA from the fragment of interest in a twofold pooled format, form a heteroduplex from the PCR products by heating (denaturing) and cooling (annealing), applying the endonuclease enzyme CEL I to digest mismatches such as SNPs and INDELS in the heteroduplex, and detect any digested fragments by separation on a LI-COR 4300 DNA analysis system. A total of 45 haplotypes ranging from simple to complex were observed from ten primer sets (Table 8.6). The mean number of SNPs and INDELS detected per marker was 13.1 and 2.6,

Table 8.6 Primers used for EcoTILLING of mungbean

Name	Target	Forward	Reverse	$MgCl_2$ (mM)	AT_m (°C)
BTF3b	IS	TCAAAAGTCTCCCCGGGGACAAGA	CCAAAGTACAAGCATCTATTGCTGCCA	4.50	61
CDC2	IS	CAACTTTGCAAGGGTGTTGCTTTCT	ACTAACACCTGGCCACACATCTTCA	4.25	65
BP1	Unknown	GTTATGGAGTTGATGAGAGGTGTCAGATA	TTGGTAAGTTCTGGAAAATGCCAACCATA	3.75	65
AIGP	IS	CTGATAGGGCCAGGAGGCAGGGAAGA	GTTTTTTAGCATTTGGACGAATGGTTGGT	3.75	60
ATCP	IS	AACCAATTGGTATTGCAGCTCAGAGCCA	TTCCTTGCCAAGAACAAACCGAATGTCA	3.75	65
CALTL	IS	GTGGAAGGCACCATTGATTGACAAC	TCTTCTTCTCAGCCTCTTCAAATGC	3.75	67
MSU380	IS	CACTCATTGCAATTTCCATGCTTCA	CAGTTGTTGTAGCAAGGGCACA	3.75	65
RL3B	Unknown	GACACGGTTCTTTGGGATTTCTC	CCTGGCTTTTCGACTTCTCTGAC	3.75	63
DNABP	IS	CAAGACATGGCTCCAATGAG	AAGAGGTAGGCGCTTTTGTG	3.00	65
SHMT	IS	CCAAACAAGGAAAAGAGGTAA	TGACTTATTCACCCCATCCA	4.25	55

Reprinted from Barkley NA, Wang ML, Gillaspie AG, Dean RE, Pederson GA and Jenkins TM (2008) Discovering and verifying DNA polymorphisms in a mung bean [*V. radiata* (L.) Wilczek] collection by EcoTILLING and sequencing. BMC Res Notes 1:28. With permission from BioMed Central, Inc.
IS intron spanning

respectively. Overall, 157 DNA polymorphisms were detected when comparing *V. radiata* var. *sublobata* and *V. radiata* var. *radiata* with a mean of 15.7 polymorphisms per marker. A total of 52 SNPs were identified and no INDELS were observed among the *V. radiata* var. *radiata* pooled accessions.

Cross Species Transferability of Greengram Microsatellite Markers

Cross-species amplification of the 127 microsatellite markers was assessed in 24 taxa of legumes in the tribe *Phaseoleae* including genus *Vigna* (African and Asian *Vigna*), *Phaseolus* and *Glycine*. One hundred and twenty five primer pairs successfully amplified DNA from more than one legume. Five primer pairs were able to amplify DNA of all legume taxa tested; while VR339 amplified only one legume species, *V. aconitifolia* (Tangphatsornruang et al. 2009). In most cases, greengram microsatellite primers were able to amplify DNA of other *Vigna* species. The transferability rates of greengram primers were between 45.80 % (*V. subterranean*) and 91.60 % (*V. angularis*). However, the amplification rate was reduced in *Phaseolus vulgaris* and *Glycine max* to 22.90 % and 24.43 %, respectively. Transferability rate of greengram genomic microsatellite markers to other *Vigna* species appeared to be more or less similar in various studies. Somta et al. (2009) reported that amplification of genic microsatellite markers in 19 taxa of *Vigna* species was between 80 % (*V. aconitifolia*) and 95.3 % (*V. reflex-pilosa*). Whereas, Chaitieng et al. (2006) reported that the amplification of azuki bean (*V. angularis*) microsatellite markers in *V. mungo*, *V. radiata*, *V. aconitifolia* and *V. umbellata* was between 68.8 and 90.2 %. The high amplification rates of both greengram and azuki bean microsatellite markers in *Vigna* species indicate high genome homology among species in this genus and are useful for genetics and genomics studies, especially genome mapping and comparative genomics.

Conclusion

The production and productivity of greengram and blackgram is limited by a combination of biotic and abiotic constraints. With many genomic tools and resources for legumes becoming increasingly available, a more detailed and in-depth genome mapping of greengram and blackgram is crucial for their genetic improvement. The current genetic linkage maps of greengram and blackgram display an inadequate level of marker density. To improve the utility of such maps, it will be required to further saturate the map with additional markers. High degree of colinearity and conservation in genome organization among legume species can be exploited for cross species utilization of identified marker/genes/DNA sequence from other legume species. For example, SSRs from azuki bean, common bean and cowpea will be useful in development of greengram linkage map with 11 LGs resolved, as

in the case of blackgram. Moreover, the information obtained from sequencing of soybean genome, common bean ESTs, and genespace of cowpea, can create high-throughput genetic markers for greengram and blackgram. In addition, a database of thousands of cowpea genespace sequences containing SSRs is now publicly available. *In-silico* development of cowpea SSRs and application of those markers in greengram and blackgram is also interesting. The partial genomic and EST sequence information available for greengram can be used for developing new markers to saturate linkage maps in greengram and blackgram. Along with the refining of linkage map, the development of physical maps linking genetically defined markers with DNA fragments is essential for the future map based cloning of genes. The BAC library of greengram has already been constructed which provides 3.5× coverage of the genome (Miyagi et al. 2004). This library has been successfully used to develop PCR based markers linked closely with a major locus conditioning bruchid resistance. Similar efforts on developing BAC library need to be attempted in blackgram. Although some progress in genome sequencing has been made in greengram, it is still far behind the other major legume crops such as soybean, cowpea, and common bean, or even their relative but less important, azuki bean. Next generation sequencing platforms have already made their strides in greengram sequencing and complete sequences of chloroplast, mitochondria, partial genome sequence and transcriptomic resources are now available in public domain. This coupled with complete genome sequencing in greengram and blackgram can generate large scale SNPs, SSRs and intron length polymorphic markers, which can help to saturate the linkage maps. They are expected to enhance molecular breeding such as marker assisted backcrossing and marker-assisted recurrent selection. This will also be helpful in development of climate resilient cultivars in the present context of climate change, resistant to serious insects, diseases and with tolerance to adverse environmental conditions. This will lead to enhanced crop productivity in these crops and ensure progress towards attaining nutritional security.

References

Afzal MA, Haque MM, Shanmugasundaram S (2004) Random amplified polymorphic DNA (RAPD) analysis of selected mungbean (*Vigna radiata* (L.) Wilczek) cultivars. Asian J Plant Sci 3:20–24

Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. Theor Appl Genet 114:359–372

Ajibade SR, Weeden NF, Chite SM (2000) Inter simple sequence repeat analysis of genetic relationships in the genus *Vigna*. Euphytica 111:47–55

Alverson AJ, Zhuo S, Rice DW, Sloan DB, Palmer JD (2011) The Mitochondrial Genome of the Legume *Vigna radiata* and the Analysis of Recombination across Short Mitochondrial Repeats. PLoS One 6:e16404. doi:10.1371/journal.pone.0016404

Applebaum SW, Marco S, Birk Y (1969) Saponins as possible factors of resistance of legume seeds to the attack of insects. J Agric Food Chem 17:618–622

Arumuganathan K, Earle DE (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

AVRDC (1984) AVRDC progress report. Asian vegetable research and development center, Shanhua, Taiwan, Republic of China

Barkley NA, Wang ML, Gillaspie AG, Dean RE, Pederson GA, Jenkins TM (2008) Discovering and verifying DNA polymorphisms in a mung bean [*V. radiata* (L.) Wilczek] collection by EcoTILLING and sequencing. BMC Res Notes 1:28

Basak J, Kundagrami S, Ghose TK, Pal A (2004) Development of Yellow mosaic virus (YMV) resistance linked DNA marker in *Vigna mungo* from populations segregating for YMV reaction. Mol Breed 14:375–383

Betal S, Chowdhury PR, Kundu S, Raychaudhuri SS (2004) Estimation of genetic variability of *Vigna radiata* cultivar by RAPD analysis. Biologia Plantarum 48:205–209

Bhat KV, Lakhanpaul S, Chadha S (2005) Amplified fragment length polymorphism (AFLP) analysis of genetic diversity in Indian mungbean (*Vigna radiata* (L.) Wilczek) cultivars. Indian J Biotech 4:56–64

Boutin SR, Young ND, Olson TC, Yu ZH, Shoemaker RC, Vallejos CE (1995) Genome conservation among three legume genera detected with DNA markers. Genome 38:928–937

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847–854

Chaitieng B, Kaga A, Han OK, Wang XW, Wongkaew S, Laosuwan P, Tomooka N, Vaughan DA (2002) Mapping a new source of resistance to powdery mildew in mungbean. Plant Breed 121:521–525

Chaitieng B, Kaga A, Tomooka N, Isemura T, Kuroda Y, Vaughan DA (2006) Development of a black gram [*Vigna mungo* (L.)Hepper] linkage map and its comparison with an azuki bean [*Vigna angularis* (Willd.) Ohwi and Ohashi] linkage map. Theor Appl Genet 113:1261–1269

Chankaew S, Somta P, Sorajjapinun W, Srinives P (2011) Quantitative trait loci mapping of *Cercospora* leaf spot resistance in mungbean, *Vigna radiata* (L.) Wilczek. Mol Breed 28:255–264

Chen HM, Liu CA, Kuo CG, Chien CM, Sun HC, Huang CC, Lin YC, Ku HM (2007) Development of a molecular marker for a bruchid (*Callosobruchus chinensis* L.) resistance gene in mungbean. Euphytica 157:113–122

Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB et al (2004) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci U S A 101:15289–15294

Choudhary S, Sethy NK, Shokeen B, Bhatia S (2009) Development of chickpea EST-SSR markers and analysis of allelic variation across related species. Theor Appl Genet 118:591–608

Chupp C (1953) A monograph of the fungus genus *Cercospora*. Cornell University, Ithaca, NY

Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Phil Trans R Soc B 363:557–572

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676

Dan Z, Cheng XZ, Wang LX, Wang SH, Ma YL (2010) Construction of Mungbean genetic linkage map. Acta Agron Sin 36:932–939

Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC, Mian MAR (2004) *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. Theor Appl Genet 108:414–422

Fatokun CA, Menancio-Hautea DI, Danesh D, Young ND (1992) Evidence for Orthologous Seed Weight Genes in Cowpea and Mung Bean Based on RFLP Mapping. Genetics 132:841–846

Gupta SK, Bansal R, Gopalakrishna T (2012) Development of intron length polymorphism markers in cowpea [*Vigna unguiculata* (L.) Walp.] and their transferability to other *Vigna* species. Mol Breed 30:1363–1370. doi:10.1007/s11032-012-9722-y (online first)

Gupta SK, Gopalakrishna T (2009) Genetic diversity analysis in blackgram (*Vigna mungo* (L.) Hepper) using AFLP and transferable microsatellite markers from azuki bean (*Vigna angularis* (Willd.) Ohwi & Ohashi). Genome 52:120–128

Gupta SK, Gopalakrishna T (2010) Development of unigene-derived SSR markers in cowpea (*Vigna unguiculata*) and their transferability to other *Vigna* species. Genome 53:508–523

Gupta SK, Souframanien J, Gopalakrishna T (2008) Construction of a genetic linkage map of black gram, *Vigna mungo* (L.) Hepper, based on molecular markers and comparative studies. Genome 51:628–637

Gupta S (2012) Project Coordinator's Report. All India coordinated research project on MULLaRP, Kanpur, India

Gwag JG, Chung JW, Chung HK, Lee JH, Ma KH, Dixit A, Park YJ, Cho EG, Kim TS, Lee SH (2006) Characterization of new microsatellite markers in mung bean, *Vigna radiata* (L.). Mol Ecol Resour 6:1132–1134

Han OK, Kaga A, Isemura T, Wang XW, Tomooka N, Vaughan DA (2005) A genetic linkage map for azuki bean [*Vigna angularis* (Willd.) Ohwi & Ohashi]. Theor Appl Genet 111:1278–1287

Hanai LR, de Campos T, Camargo LEA, Benchimol LL, de Souza AP, Melotto M et al (2007) Development, characterization, and comparative analysis of polymorphism at common bean SSR loci isolated from genic and genomic sources. Genome 50:266–277

Hisano H, Sato S, Isobe S, Sasamoto S, Wada T, Matsuno A et al (2007) Characterization of the soybean genome using EST-derived microsatellite markers. DNA Res 14:271–281

Humphry M, Konduri V, Lambrides C, Magner T, McIntyre C, Aitken E, Liu C (2002) Development of a mungbean (*Vigna radiata*) RFLP linkage map and its comparison with lablab (*Lablab purpureus*) reveals a high level of colinearity between the two genomes. Theor Appl Genet 105:160–166

Humphry ME, Lambrides CJ, Liu CJ (2005) Relationships between hard-seededness and seed weight in mungbean (*Vigna radiata)* assessed by QTL analysis. Plant Breed 124:292–298

Humphry ME, Magner T, McIntyre CL, Aitken EAB, Liu CJ (2003) Identification of a major locus conferring resistance to powdery mildew (*Erysiphe polygoni* DC) in mungbean (*Vigna radiata* L. Wilczek) by QTL analysis. Genome 46:738–744

Imrie BC, Lawn RJ, Yeates S (1991) Mungbean breeding and cultivar development in Australia. In: B.C. Imrie and R.J. Lawn (eds), Mungbean: the Australian experience. Pro. 1st Australian Mungbean Workshop, CSIRO Division of Tropical Crops and Pastures, Brisbane, Australia, pp. 14–20

Kaga A, Ishimoto M (1998) Genetic localization of a bruchid resistance gene and its relationship to insecticidal cyclopeptide alkaloids, the vignatic acids, in mungbean (*Vigna radiata* L. Wilczek). Mol Gen Genet 258:378–384

Kaga A, Ishii T, Tsukimoto K, Tokoro E, Kamijima O (2000) Comparative molecular mapping in *Ceratotropis* species using an interspecific cross between azuki bean (*Vigna angularis*) and rice bean (*V. umbellata*). Theor Appl Genet 100:207–213

Kaga A, Ohinishi M, Ishili T, Kamijima O (1996) A genetic linkage map of azukibean constructed with molecular and morphological markers using an interspecific population (*V. angularis x V. nakashimae*). Theor Appl Genet 93:658–663

Kasettranan W, Somta P, Srinives P (2010) Mapping of quantitative trait loci controlling powdery mildew resistance in Mungbean (*Vigna radiata* (L.) Wilczek). J Crop Sci Biotech 3:155–161

Keim P, Diers BW, Shoemaker RC (1990) Genetic analysis of soybean hard seededness with molecular markers. Theor Appl Genet 79:465–469

Kumar SV, Tan SG, Quah SC, Yusoff K (2002a) Isolation and characterization of seven tetranucleotide microsatellite loci in mungbean, *Vigna radiata*. Mol Ecol Notes 2:293–295

Kumar SV, Tan SG, Quah SC, Yusoff K (2002b) Isolation of microsatellite markers in mungbean, *Vigna radiata*. Mol Ecol Notes 2:96–98

Lakhanpaul S, Chadha S, Bhat KV (2000) Random amplified polymorphic DNA (RAPD) analysis in Indian mungbean (*Vigna radiata* (L.) Wilczek) cultivars. Genetica 109:227–234

Lambrides CJ (1996) Breeding for improved seed quality traits in mungbean (*Vigna radiata* L. Wilczek) using DNA markers. Ph.D. dissertation, University of Queensland, Brisbane, Australia

Lambrides CJ, Imrie BC (2000) Susceptibility of mungbean varieties to the bruchid species *Callosobruchus maculatus* (F.), *C. phaseoli* (Gyll.), *C. chinensis* (L.) and *Acanthoscelides obtectus* (Say.) (Coleoptera: Chrysomelidae). Aust J Agric Res 51:85–89

Lambrides CJ, Godwin ID, Lawn RJ, Imrie BC (2004) Segregation distortion for seed test a colour in mungbean (*Vigna radiata* L. Wilczek). J Hered 95:532–535

Lambrides CJ, Lawn RJ, Godwin ID, Manners J, Imrie BC (2000) Two genetic linkage maps of mungbean using RFLP and RAPD markers. Aust J Agric Res 51:415–425

Lawn RJ, Williams RW and Imrie BC (1988) Potential of wild germplasm as a source of tolerance to environmental stresses in mungbean. In: S. Shanmugasundaram and B.T. McLean (eds), Proc. 2nd Intl. Symp. on Mungbean, Taiwan: AVRDC. pp. 136–145

Lee JM, Grant D, Vallejos CE, Shoemaker RC (2001) Genome organization in dicots. II. *Arabidopsis* as a 'bridging species' to resolve genome evolution events among legumes. Theor Appl Genet 103:765–773

Lei S, Cheng XZ, Wang SH, Wang LX, Liu CY, Mei L, Xu N (2008) Heredity analysis and gene mapping of bruchid resistance of a mungbean cultivar V2709. Agricul Sci China 7:672–677

Maiti S, Basak J, Kundagrami S, Kundu A, Pal A (2011) Molecular marker-assisted genotyping of mungbean yellow mosaic India virus resistant germplasms of mungbean and urdbean. Mol Biotech 47:95–104

Malik IA (1996) Mungbean breeding. Lecture handout of the 14th Regional Training Course in Vegetable Production and Research, 15 October–15 March 1996, Kasetsart University, KamphaengSaen, Nakhon Pathom. Thailand: ARC-AVRDC. pp. 1–7

McConnell M, Mamidi S, Lee R et al (2010) Syntenic relationships among legumes revealed using a gene-based genetic linkage map of common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 121:1103–1116

Menancio-Hautea D, Kumar L, Danesh D, Young ND (1993) RFLP linkage map for mungbean (*Vigna radiata* (L.) Wilczek). In: O'Brien SJ (ed) Genetic maps. Cold Spring Harbor Press, New York, NY, pp 6.259–6.260

Miyagi M, Humphry M, Ma ZY, Lambrides CJ, Bateson M, Liu CJ (2004) Construction of bacterial artificial chromosome libraries and their application in developing PCR-based markers closely linked to a major locus conditioning bruchid resistance in mungbean (Vigna radiata L. Wilczek). Theor Appl Genet 110:151–156

Moe KT, Chung JW, Cho Y, Moon JK, Ku JH, Jung JK, Lee J, Park YJ (2011) Sequence information on simple sequence repeats and single nucleotide polymorphisms through transcriptome analysis of mungbean. J Integr Plant Biol 53:67–73

Narasimhan A, Patil BR, Datta S, Kaashyap M (2010) Genetic diversity assessment across different genotypes of mungbean and urdbean using molecular markers. Electron J Plant Breed 1:379–383

Peakall R, Gilmore S, Keys W, Morgante M, Rafalski A (1998) Cross amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. Mol Biol Evol 15:1275–1287

Ranade R, Gopalakrishna T (2001) Characterization of blackgram (*Vigna mungo* (L.) Hepper) varieties using RAPD. Plant Var Seeds 14:227–233

Reddy KS, Pawar SE, Bhatia CR (1994) Inheritance of powdery mildew (*Erysiphe polygoni* DC.) resistance in mungbean (*Vigna radiata* L. Wilczek). Theor Appl Genet 88:945–948

Saini A, Reddy SK, Jawali N (2004) Evaluation of long primers for AP-PCR analysis of mungbean (*Vigna radiata* (L.) Wilczek): genetic relationships and fingerprinting of some genotypes. Indian J Biotechnol 3:511–518

Santalla M, Power JB, Davey MR (1998) Genetic diversity in mungbean germplasm revealed by RAPD markers. Plant Breed 117:473–478

Sarkar S, Swagata G, Mitali C, Padminee D, Triparna L, Anirban M, Nirmal M, Kanti P, Somnath B (2011) Molecular markers linked with bruchid resistance in *Vigna radiata* var. s*ublobata* and their validation. J Plant Biochem Biotechnol 20:155–160

Selvi R, Muthiah AR, Manivannan N, Raveendran TS, Manickam A, Samiyappan R (2006) Tagging of RAPD for MYMV resistance in mungbean (*Vigna radiata* (L.) Wilczek). Asian J Plant Sci 5:227–280

Sholihin HDM (2002) Molecular mapping of drought resistance in Mungbean (*Vigna radiata*): 1. Linkage map in Mungbean using AFLP markers. J Bioteknologi Pertanina 7:17–24

Singh S (2003) DNA markers in mungbean. Ph.D. thesis, University of Mumbai, India

Singh S, Reddy KS, Jawali N (2000) PCR analysis of mungbean genotypes using anchored simple sequence repeat primers. In: DAE-BRNS symp. on the use of nuclear and molecular techniques in crop Improvement. Mumbai: BARC. pp. 359–369

Sivaprakash KR, Prashanth SR, Mohanty BP, Parida A (2004) Genetic diversity of blackgram (*Vigna mungo*) landraces as evaluated by amplified fragment length polymorphism markers. Curr Sci 86:1411–1416

Somta P, Seehalak W, Srinives P (2009) Development, characterization and cross-species amplification of mungbean (*Vigna radiata*) genic microsatellite markers. Conserv Genet 10:1939–1943

Souframanien J, Gopalakrishna T (2004) A comparative analysis of genetic diversity in black gram using RAPD and ISSR markers. Theor Appl Genet 109:1687–1693

Souframanien J, Gopalakrishna T (2006) ISSR and SCAR markers linked to the mungbean yellow mosaic virus (MYMV) resistance gene in blackgram [*Vigna mungo* (L.) Hepper]. Plant Breed 125:619–622

Souframanien J, Gopalakrishna T (2009) Cross-species amplification of microsatellite loci and diversity analyses in blackgram. J Food Legumes 22:11–17

Souframanien J, Gupta SK, Gopalakrishna T (2010) Identification of quantitative trait loci for bruchid (*Callosobruchus maculatus*) resistance in black gram [*Vigna mungo* (L.) Hepper]. Euphytica 176:349–356

Souframanien J, Joshi A, Gopalakrishna T (2003) Minimal intraspecific variation in the internal transcribed spacer region of rDNA in blackgram (*Vigna mungo* (L.) Hepper). Curr Sci 85:798–802

Souframanien J, Pawar SE, Rucha AG (2002) Genetic variation in gamma ray induced mutants in blackgram as revealed by RAPD and ISSR markers. Indian J Genet Plant Breed 62:291–295

Tangphatsornruang S, Sangsrakru D, Chanprasert J, Uthaipaisanwong P, Yoocha T, Jomchai N, Tragoonrung S (2010) The Chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. DNA Res 17:11–22

Tangphatsornruang S, Somta P, Uthaipaisanwong P, Chanprasert J, Sangsrakru D, Seehalak W, Sommanas W, Tragoonrung S, Srinives P (2009) Characterization of microsatellites and gene contents from genome shotgun sequences of mungbean (*Vigna radiata* (L.) Wilczek). BMC Plant Biol 9:137

Timmerman-Vaughan GM, McCallum JA, Frew TJ, Weeden NF, Russell AC (1996) Linkage mapping of quantitative trait loci controlling seed weight in pea (*Pisum sativum* L.). Theor Appl Genet 93:431–439

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55

Williams RW (1989) A study of the causes of, and selection for resistance to weather damage in mungbean (*Vigna radiata* (L.) Wilczek; *V. mungo* (L.) Hepper). Ph.D. dissertation, University of Queensland, Brisbane, Australia

Young ND, Danesh D, Menecio-Hautea D, Kumar L (1993) Mapping Oligogenic resistance to powdery mildew in mungbean with RFLPs. Theor Appl Genet 87:243–249

Young ND, Kumar L, Menancio-Hautea D, Danesh D, Talekar NS, Shanmugasundaram S, Kim DH (1992) RFLP mapping of a major bruchid resistance gene in mungbean (*Vigna radiata* L. Wilczek). Theor Appl Genet 84:839–844

Yu K, Park SJ, Poysa V (1999) Abundance and variation of microsatellite DNA sequences in beans (*Phaseolus* and *Vigna*). Genome 42:27–34

Zhang MC, Wang DM, Zheng Z, Humphry M, Liu CJ (2008) Development of PCR-based markers for a major locus conferring powdery mildew resistance in mungbean (*Vigna radiata*). Plant Breed 127:429–432

Chapter 9
Common Bean Genomics and Its Applications in Breeding Programs

Juan M. Osorno and Phillip E. McClean

Abstract Because of its nutritional value, easiness of cultivation, and cultural preference in many cases, common bean (*Phaseolus vulgaris* L.) is the most important grain legume in the human diet worldwide. Recent genomic evidence suggest that common bean originated in Central America and confirms the two centers of domestication previously characterized (Mesoamerican and Andean), with well-defined races within each gene pool. Total world production of dry bean from the 10 year period 1961–1970 increased 65 % to 169 million MT in the period 2001–2010. The main challenge now is how to apply these genomic tools into breeding programs for increased efficiency. Applications go from marker-assisted breeding to tracking of F_1 crosses, and even DNA fingerprinting, among others. More recently, the development of thousands of single nucleotide polymorphisms (SNP) markers and the completion of the bean genome sequence have opened numerous opportunities for fine mapping and gene characterization. The exploitation of linkage disequilibrium through association mapping allows for rapid identification of important genomic regions associated with traits of economic importance without the need of creating bi-parental populations for this goal. The following sections will describe specific examples of applications of these genomic tools into breeding programs and illustrate some of the possible future directions some of these technologies may follow.

Keywords *Phaseolus vulgaris* • Disease resistance • Genomic resources • Gene pools • Whole genome sequencing • Molecular markers • Marker assisted selection • MAGIC population • Single nucleotide polymorphism • Genotyping by sequencing

J.M. Osorno, Ph.D. (✉) • P.E. McClean, Ph.D.
Department of Plant Sciences, North Dakota State University,
NDSU Dept. 7670, P.O. Box 6050, Fargo, ND 58108-6050, USA
e-mail: juan.osorno@ndsu.edu

S. Gupta et al. (eds.), *Legumes in the Omic Era*, DOI 10.1007/978-1-4614-8370-0_9,

Introduction

Common bean (*Phaseolus vulgaris* L.) is the most important grain legume in the human diet and has been described as a nutritional powerhouse (Broughton et al. 2003; Miklas and Singh 2007; Câmara et al. 2013). Common beans are used in a range of ways, but dry beans are the most frequently found because of their simple storage and shell life. However, beans can be also consumed as tender pods (snap or green beans) and fresh (threshed manually at physiological maturity). Recent genomic evidences suggest that common bean originated in Central America (Papa et al. 2007; Bitocchi et al. 2012) and has two centers of domestication (Mesoamerican and Andean) with well-defined races within each gene pool (Singh et al. 1991). Common beans were widely cultivated in Mexico and the U.S. during pre-Columbian times. New World settlers cultivated dry beans from European introductions in the eastern U.S. and from landraces of small red, pink, pinto and great northern beans in the western U.S. that were cultivated by Native Americans.

The *Phaseolus* genus has five species that have been domesticated namely, *P. vulgaris* or common bean; *P. polyanthus* or year-long bean; *P. coccineus* or scarlet runner bean; *P. acutifolius* or tepary bean; and *P. lunatus* or lima bean (Freytag and Debouck 2002). Genetic diversity is wide across market classes but narrow within each class because of intensive crosses mostly among elite lines (Acosta-Gallegos et al. 2007). Total world production of dry bean from the 10 year period 1961–1970 increased 65 % to 169 million MT in the period 2001–2010 (USDA-ERS 2013). During this same period, U.S., dry bean production increased 71.0 % to 11.5 million MT and accounted for 5.8 % of world production. Statistics from the Food and Agriculture Organization (http://faostat3.fao.org/home/index.html) list India, Myanmar, Brazil, China, and United States as the top five producers. Unfortunately, this database pools together both *Phaseolus* beans and species now belonging to the *Vigna* genera (e.g. golden, green, and black gram, mung beans, adzuki beans, among others). Therefore, it is difficult to obtain reliable estimates of production of *Phaseolus* beans from some of these countries (Beebe 2012). Nonetheless, it is well known that most dry bean production in India, Myanmar, and China is focused mainly on mung beans and adzuki beans because of a cultural preference. Contrastingly, *Phaseolus* beans are the preferred in countries such as Brazil and United States. As a matter of fact, *Phaseolus* beans are the most important edible legume in all the American continent, Africa, and Europe.

Genomic research has greatly expanded the resources available for plant breeders. Principal among these are the vast array of molecular markers that breeders can use for many aspects of their breeding program. At the earliest stages, the simplest application is confirming that a cross between two parents was successful by screening the F_1 progenies. This requires just a single marker that is polymorphic between two parents. Later in a program, a breeder may simple wish to follow a few traits that are essential for the target region. Most often, this will be a marker for essential disease resistance genes or other simply inherited traits (Miklas et al. 2006). At the end of the breeding process, a line near commercial release might be fingerprinted

with a suite of marker loci for identity preservation purposes or any possible challenges regarding the ownership of the germplasm. These marker assays are generally straight forward, and in most cases, they are affordable for most small to medium-sized breeding programs. Breeders are now being asked to not only monitor their programs, but also using their programs for marker/trait association studies. Given that many of the simple traits have abundant markers available, the traits of interest now are complex or quantitative in nature with heritability estimates that range from low to high. To dissect the genetic effects controlling these traits, a much larger collection of markers is necessary. Today, most crop species have developed single nucleotide polymorphism (SNP) marker systems. Since the nucleotide is most basic element of the DNA code, this marker system provides the greatest level of granularity when look for causative or associated markers. SNP markers are one of the major outcomes from the development of genomic resources for a crop, and they that are facilitating new directions in plant breeding and genetics.

Common bean breeders have been probably one of the earliest adopters of molecular markers among the crops of economic importance (Beaver and Osorno 2009). Isozymes and seed storage proteins (SDS-PAGE) were initially used to characterize the genetic diversity of the *Phaseolus* species complex. However, the big advances were made when DNA-based markers became available. Marker systems such as RFLPs allowed obtaining a lot of specific information about each genotype, but the process was cumbersome and somehow dangerous. Nonetheless, the first linkage maps for common bean were obtained and allowed a better understanding of the genomic structure of common bean (Gepts et al. 2008).

It was not until the development of the polymerase chain reaction (PCR) during the early 1990s that DNA-based markers became a reality in many laboratories and more importantly, within breeding programs. In the specific case of common beans, RAPD markers were intensively used not only for to characterize genetic diversity, but to improve the linkage maps initially assembled. Up today, many of these RAPD markers that were positively associated with traits of interest have been transformed into SCARs in order to make them more stable, reliable, and easier to score (Miklas et al. 2006). This allowed a routine use of these SCAR markers within bean breeding programs as another selection tool, especially for major genes associated with disease resistance (http://bic.css.msu.edu/_pdf/SCAR_Markers_2010.pdf).

Microsatellite markers (SSR) have been intensively used in gene mapping and genetic diversity studies since the mid 1990s (Blair et al. 2003). When many of these marker systems were used together to assemble genetic maps and saturate specific regions of the genome, it allowed to narrow down genomic regions, tag specific genes, obtain their sequences (EST sequencing), and finally, identify and clone candidate genes (Blair et al. 2011). Genomic and cDNA clones have been useful for marker development and gene mining, and they form the basis for some of the recent sequencing projects. These include an expressed sequence tag (EST) effort consisting of 22,000 sequences derived from four cDNA libraries made from the Mesoamerican genotype Negro Jamapa, and one cDNA library made from the Andean genotype, G19833 (Ramirez et al. 2005).

More recently, the development of thousands of Single Nucleotide Polymorphisms (SNP) and the completion of the bean genome have opened numerous opportunities for fine mapping and gene characterization (Hyten et al. 2010). The exploitation of linkage disequilibrium through association mapping allows for rapid identification of important genomic regions associated with traits of economic importance without the need of creating bi-parental populations for this goal. The idea of simultaneous breeding and gene mapping is now possible thanks to genomic selection and genotyping by sequencing methods. The following sections will illustrate specific examples of applications of these genomic tools into common bean breeding programs and illustrate some of the future directions some of these technologies may follow.

Marker Assisted Selection

Cultivar development is complex because breeders work with multiple traits simultaneously and attempt to bring them together into the same genotype. This is even more challenging when the genetic nature of each trait and the interaction with the environment are taken into consideration. In spite of these challenges, it is hard to deny the progress accomplished especially during the last century. Some crops have seen larger yield increases than others and in many cases they are reaching a yield plateau, even in crops in which scientific resources are abundant (Duvick and Cassman 1999; Evans and Fisher 1999; Ray et al. 2012). For some of those crops, genomic tools have been in part responsible for some of those yield increases especially during the last 20 years, but more is needed. In the specific case of common bean, yield gains have not been as steep as in other crops (Vandemark et al. 2013). Nonetheless, the development of several markers linked to disease resistance genes have allowed to follow these genomic regions across the breeding/selection process (Miklas et al. 2006). Similar to other crops, successful stories are more frequently found with major genes while marker-assisted selection for quantitative traits have been more challenging.

The idea of selecting a suite of desirable traits from a given genomic landscape is today a more real possibility for many crops including common beans. This will allow for more efficient phenotyping efforts by reducing population sizes at early generations and increasing the probabilities of finding the optimum recombinant genotype. However, the success of these methodologies will be affected by several other factors such as the number of traits under selection, their genetic control and their interaction with other traits and the environment.

Co-evolution of host and pathogen has led to isolates and races of Andean origin which attack beans primarily from the Andean gene pool. Conversely, isolates of Middle American origin attack beans primarily in the Middle American gene pool but possess a wider range of virulence also infecting beans of Andean origin. Examples of this can be found in angular leaf spot (Guzmán et al. 1995), anthracnose (Balardin and Kelly 1998), common bacterial blight (Mkandawire et al. 2004)

and rust (Sandlin et al. 1999). Co-evolution of pathogen virulence within gene pools plays a key role when designing breeding strategies. Resistance genes of Middle American origin are very effective when transferred to beans of Andean background and deployed in regions where Andean isolates prevail (East Africa and South America). Similarly, genes of Andean origin are very effective when transferred to beans of Middle American background and deployed in regions where isolates of Middle American origin prevail (Central America, Mexico, USA). In addition, similar issues are found when a marker works efficiently in one gene pool but not in the other, and vice versa.

Since 2000, over 90 % of the disease-resistant germplasm and few cultivars released in the US were developed using molecular markers. This is a good first start but is limited by the fact that many of these markers are at several cM from the actual gene controlling the trait. Nonetheless, the best marker for a particular trait is the gene that controls the phenotype. Second best, would be a marker that essentially co-segregates with the functional gene by mapping very close to it. The best approach in either case is to genetically map a trait to a genomic interval and use a candidate gene approach that combines gene model data with best estimates regarding the type of gene that would control that trait. Some of the best information regarding function is from the field of plant pathology where consensus regarding the major structure of disease resistance genes has evolved.

As mentioned previously, there are interesting and early applications of genomic tools in common bean breeding programs, ranging from using seed storage proteins such as *Phaseolin*, to selection for agronomic and seed nutritional components by using SNP platforms. In the case of major genes, the most successful examples can be found in genes related to disease resistance. Successful examples can be found with bean common mosaic virus (BCMV), bean golden yellow mosaic virus (BGYMV), anthracnose, common bacterial blight [*Xanthomonas campestris* pv. *phaseoli* (Smith) Dye [Syn. *X. axonopodis* pv. *phaseoli* (Smith), Vauretin et al.], and white mold (*Sclerotinia sclerotiorum* (Lib.) de Bary). Moderate success has been also observed in angular leaf spot, bean curly top virus (BCTV), bean rust (*Uromyces appendiculatus* Pers:Unger), Halo blight (*Pseudomonas syringae* pv. *phaseolicola* (Burkholder)) (Miklas et al. 2006; Beaver and Osorno 2009; Beebe 2012).

Virtually all common bean breeding programs make sure that when releasing a cultivar, it contains the *I* gene for resistance to BCMV since it is present in all bean producing regions worldwide (Beaver and Osorno 2009). The SW13 SCAR marker has proven to be a very reliable marker and is routinely used in many bean breeding programs given its accuracy, cost, and easiness of use (Miklas et al. 2006). In the case of BGYMV, the SR2 SCAR has been very useful in tropical breeding programs where this viral disease is very important.

Breeding efforts for resistance to BGYMV were possible at the International Center for Tropical Agriculture (CIAT) in Colombia by using the SR2 marker despite the fact that this virus is not present in this country. Therefore, breeding lines were initially screened with the marker and then selected lines only were sent to all different countries for further field screening (Blair et al. 2007; Beebe 2012). Even more, the first genetically modified bean has been approved for release in

Brazil and it has been engineered to avoid BGMV replication by using RNA interference (Bonfim et al. 2009). In addition, genes from common bean have been used in transformation experiments in other crops (Barbosa et al. 2010). Eapen (2008) provides a current review on the status of genetic transformation on legumes.

In the case of anthracnose, among the numerous SCAR markers available, the SB114 marker linked to *Co-4*^{2} gene has shown the highest levels of reliability (Miklas et al. 2006). Today, the *Co-4*^{2} gene is recognized as the most broad-based resistance gene and the amplification of this marker indicates a wide range of resistance genes in the germplasm accessions (Kelly and Vallejo 2004). For common bacterial blight, the BC420 and SU91 SCAR markers and respective linked QTL derive from tepary bean via breeding line XAN 159 (Miklas et al. 2000; Yu et al. 2000). Advanced cranberry, pinto, Great northern, and snap bean germplasm with combined resistance to common blight has been developed in the USA using these markers in the selection process. However, SU91 is the marker most frequently used in breeding programs (Miklas et al. 2006).

In the case of white mold, efforts are focused on two major QTLs (WM2.2 and WM8.3). SCAR markers were assayed by Soule et al. 2011 across a wide array of germplasm representing different origins and different reactions to white mold disease with reliable results. Overall, the SCARs appear less useful for marker-assisted breeding in the Andean gene pool because most lines surveyed possessed the coupling and lacked the repulsion markers regardless of disease reaction. In fact, four of them, WM2.2, WM7.1, WM7.2 and WM8.3, have already shown potential utility for marker assisted breeding (Miklas and Singh 2007; Ender et al. 2008). It is important to note that breeding strategies that combine MAS with intermittent phenotypic selection have been the most effective in developing lines with improved disease resistance. Phenotypic selection is needed to retain minor effect QTL and select for epistatic interactions that contribute to improved resistance.

Current efforts focus not only on disease resistance, but also on abiotic stresses such as drought tolerance, water use efficiency, and soil constraints. In addition, specific efforts on mapping and characterizing the genes related to seed mineral content and other nutritional traits will allow doing a more efficient improvement to increase the levels of nutritionally important elements and molecules as well as reducing the one with negative impact in nutrient availability and uptake. The existence of a genome sequence will also allow to physically map these regions more accurately, design more reliable markers, and advance towards candidate-gene approaches faster.

Common Bean Genome Sequence and Database Development

The genesis of the United States common bean sequence project was a white paper shared with multiple funding agencies (McClean et al. 2008). This document outlined a strategy to use next-generation sequence technology, that was at the time replacing the then standard Sanger sequencing approach, for the bulk of the sequencing, while Sanger sequencing was limited to a bacterial artificial chromosome (BAC)-end

sequence survey. In early 2009, the United States Department of Agriculture (USDA) requested projects for sequencing of the genome, and funding was obtained for the project in September of that year. The investigators on the project were Scott Jackson (PI, Univ. of Georgia), Phillip McClean (North Dakota State Univ.), Jeremy Schmutz (HudsonAlpha Institute for Biotechnology), Daniel Rokhsar [Department of Energy, Joint Genome Institute, (DOE/JGI)], and Perry Cregan (USDA/ARS). In addition, funds were secured by the DOE/JGI to also support the project.

Because of the availability of a public BAC library (Schlueter et al. 2008; Córdoba et al. 2010), the Andean genotype G19833 was chosen as the reference genome for the project. The bulk of the sequence data was collected using the Roche 454 technology in the form of ~400 bp reads. This data was augmented with BAC-end and fosmid Sanger sequence. Following the development of an initial set of contigs and scaffolds, pseudochromosomes were assembled by alignment of the contigs and scaffolds to a new SNP map consisting of ~7,000 SNP loci ordered using an F_2 population from the cross of the pinto cultivar Stampede and the dark red kidney cultivar Redhawk. The SNPs were primarily developed the USDA funded Common Bean Coordinated Agricultural Project (BeanCAP: http://www.beancap.org). To assist with gene modeling and annotation, RNA-seq data was collected from an array of tissues (root, stem, flower, and pod) using Illumina technology.

The common bean genome sequence was recently released (August 2012). The assembled genome size is 521 Mb and consists of 41 % repetitive DNA. Much of the repetitive DNA is relatively new in origin, and many of the genes exist in repeat islands. 27,197 gene models were defined, and 4,441 alternative splice variants were discovered for a total of 31,638 protein-coding transcripts. The genome sequence is fully accessible (in pre-publication stage), at the DOE/JGI Phytozome website (http://www.phytozome.net/search.php?method=Org_Pvulgaris). Here all the gene models and associated sequences and Pfam domain data can be found. The sequence is also available in the Legume Information System (LIS: http://phavu.comparative-legumes.org/gb2/gbrowse/Pv1.0/). This database is dedicated to a comparative analysis of legumes. Given the ancestral relatedness between common bean, soybean, and pigeon pea (*Cajanus cajan* L. Millsp.), the synteny blocks tracks offered by this site will be useful in future research as breeders like to mine phenotypic syntenies between species to discover shared causative genes for important agronomic crops (Zhu et al. 2005). This shared phenotypic synteny has already been shown for the determinacy loci in common bean (*fin*: Repinski et al. 2012) and soybean (*dt1*: Tian et al. 2010) that map to common bean chromosome Pv01 and soybean chromosome Gm19. These two regions were shown to be syntenic (McClean et al. 2010).

Two additional genome sequence projects are underway at Mexico (through a multi-national effort including Brazil, Argentina, and Spain) (http://www.genomacyted.org/) and Canada (http://www.beangenomics.ca/research/projects/view/draft-genome-sequence-for-common-bean-i-p-vulgaris-i), working with BAT93 black bean and OAC-Rex navy bean, respectively. These projects will complement very well with the one described above because it will allow to have sequences from both gene pools (G18933 is Andean while the other two genotypes are Mesoamerican).

In addition, a new, breeder-friendly database, PhaseolusGenes (http://phaseolusgenes.bioinformatics.ucdavis.edu/), developed by P. Gepts (UC Davis) has collected all available and emerging marker data. Comparative genetic maps are available for more detailed marker discovery. This database is a rich resource for all of the mapped marker data for single gene and quantitative traits. It also will develop the capability to move between common bean and soybean as researchers use marker data to take advantage of phenotypic synteny (Yang et al. 2010). The utility of this database was shown as it was the key link for the recent development of markers for anthracnose (*Colletotrichum lindemuthianum* (Sacc. & Magnus) Lams.-Scrib.) and for angular leaf spot (*Phaeoisariopsis griseola* (Sacc.) Ferraris) (Gonçalves-Vidigal et al. 2011).

Markers and the Genome Sequence

As mentioned before, bean breeders are major adopters of molecular marker technology. The current marker suite used by breeders is the outcome of concerted global efforts to develop PCR-based markers associated with important disease resistance loci. These markers are publicly available and address diseases important to bean production throughout the world. The essential data for SCAR markers developed to track many important diseases is freely available (http://bic.css.msu.edu/_pdf/SCAR_Markers_2010.pdf).

The BeanCAP project has significantly increased the number of molecular markers available to over 10,000 single nucleotide polymorphism (SNPs) that are functional across many different populations. From these, a 6,000 SNP marker assay was designed for genotyping using the Illumina Infinium system. Additionally, paired-end sequence data was collected from 16 genotypes. The reads were assembled in contigs and they were searched for indels. These were then converted into a collection of 2,687 indel markers that have utility both between and within market classes and can be easily multiplexed (Moghaddam et al. 2013).

BeanCAP researchers are using the SNP platform for the discovery of markers for a number of important agronomic and seed nutritional traits. Where appropriate, that platform will also serve this project as an ideal set of markers for the bi-parental mapping experiments that are being proposed here. While a 6,000 SNP chip is useful for biparental mapping of populations with limited recombination, genotype-by-sequencing (GBS: Elshire et al. 2011) data was collected on the same BeanCAP association mapping panel. This provides a deeper assessment of the variation within a population. This amount of data also allows for a finer scale mapping and makes association mapping experiments, such as those recently performed on sorghum, more robust (Morris et al. 2013). The richer dataset is especially important given the recent results (Moghaddam et al. 2013) that shows that linkage disequilibrium in a population representing the breadth of modern bean germplasm decays to ~100 kb at $R^2 = 0.1$. While this level of LD will allow us to develop markers that reside near the gene of interest, it will require greater marker density than that

provided by 6,000 SNPs. From that perspective, the GBS data is critical. Nonetheless, the 6,000 SNP platform is already speeding up the process of gene mapping and marker identification for traits of economic importance such as the slow-darkening trait (Felicetti et al. 2012).

Alternative Mapping Approaches

During the last 35 years, the conventional bi-parental mapping approach has been very useful to tag, map and characterize numerous loci (Collard et al. 2008; Mackay and Powell 2007). However, progress has been very moderate when it comes to using those tagged regions in a routine way in breeding programs. Several reasons explain why it is difficult to apply them in a MAS scheme, but the main problem is the lack of repeatability among different genetic backgrounds, so in most cases MAS is useful when the original source of the QTL is used in the cross. The second most important limitation is the time and resources that are needed to develop a bi-parental population (e.g. F_2, RILs, NILs, etc.). In addition, many of the QTL effects initially mapped were overestimated because of small population sizes and/or few locations tested in many of the studies. Nonetheless, the more recent studies are more aware of this and are attempting to use more lines in the populations.

With the rapid development of new genomic tools, a newer wave of alternative mapping approaches has been also established. Many of these methods have been initially used in humans and animals and then adapted to plants, sometimes with different goals. In humans for example, researchers need to be more creative because of the impossibility of creating mapping populations in the same way is done in animals, insects, or plants. This is the case of genome-wide association mapping (GWAS) in which a large and diverse panel of genotypes is used to find statistical associations between traits and markers across the entire genome taking advantage of linkage disequilibrium (Rafalski 2010; Chen 2013). One of the main advantages of this mapping approach is that no crossing and population development is needed. Instead, a large set of genotypes (e.g. cultivars, germplasm, breeding lines, etc.) is assembled, accounted for population structure, and used for phenotyping and genotyping. Then several statistical methods are used to detect significant associations between markers and traits. First, the genotypes that make up the population of the study should be selected from a natural population or germplasm collection in a way that includes a wide range of phenotypic variations. In the second step, the population should be genotyped using the preferred molecular marker system and phenotyped in different environments with replicates. The genotypic information from molecular markers will be used in estimating the LD decay, population structure and coefficient of relatedness (kinship). The final step involves statistical approaches to identify the association between a phenotype and a marker locus in its proximity (Balding 2006). Genetic diversity and population structure are very important factors that can be improved by population size. In most cases, an SNP platform is used as the genotyping method and the population can be screened/evaluated for

multiple traits and environments in order to obtain reliable phenotypic information. GWAS is its initial steps in bean but it has been widely used and proven in other crops of economic importance such as maize, soybean, wheat, and rice, among others. Current efforts on beans focus on the mapping of genomic regions associated with agronomic traits, disease resistance, and nutritional traits. GWAS is very useful detecting robust QTL regions with large effects but it has not been very successful detecting QTL regions with small effects. In addition, confirmation within bi-parental populations is needed in many cases to validate results found with GWAS. Nonetheless, this is a rapid method to narrow down genomic regions of interest that can be then saturated and studies in more detail. Although most of the association studies are conducted in Arabidopsis, there are successful examples of GWAS in maize (Buckler et al. 2009; Tian et al. 2010; Yang et al. 2011), barley (*Hordeum vulgare* subsp. *vulgare*) (Caldwell et al. 2006), rice (*Oryza sativa* L.) (Zhao et al. 2011), sorghum (*Sorghum bicolor* L.) (Hamblin et al. 2005), soybean (Hyten et al. 2007; Mamidi et al. 2011) and common bean (Blair et al. 2009; Shi et al. 2011), among others.

Another type of population more recently used for gene mapping is the Multi-parent Advanced Generation Inter-Cross population or MAGIC populations (Cavanagh et al. 2008; Kover et al. 2009). It can be considered an extension of the advanced intercross in which an intercrossed mapping population is created from multiple founder genotypes, typically six to eight parents (Mackay and Powell 2007). Contrary to crossing two lines as in any bi-parental population, a MAGIC population is established by crossing together multiple founder lines or genotypes. Such populations are more genetically diverse that those established from just two parents, giving more bangs per buck: more associations can be found. In addition rather than searching for associations immediately after crossing, the population is first cycled through several additional generations of crossing. Each extra generation mills the genetic contribution from the founder lines finer and finer. As a result, associations are located with greater accuracy and are therefore of more use.

The large number of parental accessions increases the allelic and phenotypic diversity over traditional RILs, potentially increasing the number of QTL that segregate in the cross. In addition, the larger number of accumulated recombination events increase the mapping accuracy of the detected QTL compared to an F_2 cross. Thus, MAGIC populations could be considered an intermediate stage between naturally occurring accessions and existing synthetic populations. They represent a significant improvement over standard RILs descended from just two founders in that they capture more of the genetic and phenotypic variation present. Furthermore, they have a higher density of recombinants, which improves mapping resolution. However, it is important to maximize the genetic diversity in order to improve results (Huang et al. 2013).

MAGIC populations may take time to develop, but results have shown that mapping accuracy and detection is much improved in the MAGIC populations when compared to traditional bi-parental F_2 and RIL mapping populations. Moreover, the combination of MAGIC and association mapping may be beneficial. While association mapping may be able to identify QTLs with better accuracy, the population

structure observed among natural accessions requires much care to distinguish between true QTL and false positives. In comparison, the structure of the MAGIC populations is relatively simple. If there are common variants in MAGIC populations and natural accessions, the former group may provide an ideal material to verify QTL identified with association mapping. Current efforts are focused on either producing or already using these MAGIC populations in a number of crops including wheat (NIAB, CSIRO, Univ. of Bologna), oats (IBERS), rice (IRRI), barley (SAC), and sorghum (ICRISAT). Several preliminary results are being published already.

Beebe (2012) suggests that alternative population types would also be of interest for the analysis of drought resistance. In this sense, the advanced backcross strategy holds promise for the determination of QTLs that function without the confounding effect of epistasis with alleles from non-commercial sources, since advanced backcross breeding fixes valuable alleles in the genetic background of a commercial parent.

Double haploid populations have been used in mapping exercises in several other crops such as wheat, barley, maize, and rice. However, production of double haploid progenies in beans and legumes in general has been challenging. Therefore, this approach cannot be used until an efficient and reliable method for generation and development of double haploid in beans is accomplished.

Early and Future Applications of Common Bean Genomic Sequence Data

Compared to many crop species such as soybean, corn, and wheat, the common bean genome is moderate in size (~550 MBp/haploid; Bennett and Leitch. 2005). Importantly, all of the early molecular mapping experiments pointed to one important observation: common bean is a "true" diploid since nearly all marker loci map to a single location (Vallejos et al. 1992; Freyre et al. 1998; McClean et al. 2002; Gepts et al. 2005). This was recently highlighted by research findings of McClean et al. (2010) who used all available EST (expressed sequence tag) sequences to develop contig sequences that are representative of the gene space in the genome. Comparative analysis showed that while most of these were single copy in common bean, many of these genes were duplicated in soybean (*Glycine max* L.), an indication of the diploid history of common bean relative to the polyploidy history of soybean. This discovery was consistent with previous observations that the traditionally large gene families, such as resistance gene analogs (Rivkin et al. 1999) and protein kinases (Vallad et al. 2001), have fewer members in common bean than other crops.

In the early days of genomic research, EST sequences were valuable resources, primarily because they gave our first glimpse at the genes within a genome. The earliest EST project in common bean was published by Ramirez et al. (2005). This new sequence-based resource was used by McConnell et al. (2010) to begin a characterization of SNP and indel density in the genome. Sequence data was collected from 550 gene fragments from BAT93 and Jalo EEP558, the parents of a

community-wide genetic map. A diversity analysis revealed 1,580 SNPs and 130 indels, found an excess of synonymous compared to non-synonymous sites, and determined there was a slight excess of transitions than transversions.

The parents of four other popular mapping populations were also characterized. Although this analysis only compared five pairs of mapping parents, this represented the first measure of the genome-wide diversity in common bean. This data was subsequently mined to develop a suite of 300 gene-based markers distributed across the common bean genome that in turn were used to developed the first gene-based map of common bean. Such maps are now common for other plant species. These markers, termed "g markers" (for genomic markers), have now been used for mapping projects, where for example new markers for angular leaf spot (Gonçalves-Vidigal et al. 2011), common bacterial blight (Shi et al. 2011), and popping ability (Yuste-Lisbona et al. 2012), have been developed.

Other sequence-based markers systems have also been developed, including simple sequence repeats and legume-based anchor markers (Hougaard et al. 2008). These have now been electronically merged into a consensus map (Galeano et al. 2011). Although useful from a global perspective, caution is needed when using this consensus map because it is based on visual alignments of maps rather than true genetic recombination events.

Applications of Genomic Sequence Data for a Modern Breeding Program

Plant breeding programs will change with the infusion of sequence data that is now available to them. This data will allow for new approaches not utilized before for common bean improvement. Three examples are highlighted here and it describes a possible research agenda that will improve the pace of common bean improvement. The first, and probably the most obvious, is the utility of a sequenced genome to discover the actual causative gene, and then the causative mutation, either a SNP or insertion or deletion, that manifests itself as a unique phenotype of agronomic importance. Examples for common bean to follow are available from two important world-wild food crops: rice (*Oryza sativa* L.) and soybean. Rice, the first crop species with a complete reference genome (IRGSP 2005), has benefited from the sequence. For example, the gene associated with sticky rice, an essential culinary trait, was cloned using the genome sequence as a reference resource (Yamanaka et al. 2004). Yield is the major production trait for all crops, and rice researchers have discovered genes that affect seed size (Fan et al. 2006), seed width (Song et al. 2007), and seed number (Ashikari et al. 2005). With these genes in hand, molecular markers have been designed to track these important phenotypes in a breeding program (Wang et al. 2011). Soybean, the first economically important legume for which a draft genome was developed (Schmutz et al. 2010), has also benefited from that important modern genomic resource. Soybean expresses two growth habits, indeterminate and determinate. In some countries such as the United States, the

crop is grown across a wide latitudinal band. The northern region is seeded to indeterminate varieties while in the southern region determinate varieties predominate. One of the major genes that controls this phenotype, *dt1*, was recently cloned (Tian et al. 2010) utilizing the genomic resource to identify the soybean homology to a previously cloned *Arabidopsis* determinacy gene (Shannon and Meeks 1991). Using the gene sequence, it is now possible develop functional markers for this phenotype, which is tricky to score. These markers will assist breeders in maintaining the desired growth habit as they introgress important alleles for other agronomic traits from a wider germplasm basis. This will be especially important as they try to introduce resistance using sequence data from the recently cloned gene critical for soybean cyst nematode resistance (Cook et al. 2012). These are a few examples of the type of improvement approaches that are now available to common bean breeders and geneticists with the release of the genome sequence.

Association mapping trials are a second example that will greatly benefit from the genome sequence. This mapping method utilizes a large population of genotypes representing the variation found within the particular subset of the species relevant to an agronomic trait of interest (Morrell et al. 2011). For example, the BeanCAP association panel was developed to represent modern cultivars from the Middle American gene pool adapted to United States production environments. Trials such as these will become more frequent because of the ease of generating the mapping population relative to a bi-parental population. Data from these multiple trials can then analyzed using approaches currently used in human trials to discover genomic regions, and possibly the causative alleles, for traits of worldwide importance.

A human example that focused on Alzheimer's disease serves to highlight the methodology (Naj et al. 2011). A consortium was formed and phenotypic data from 17,675 patients was collected from the 29 Alzheimer Disease Centers involved in the project. All of the individuals were genotyped with one of three high-density SNP arrays, and missing data was imputed across all chips to generate a SNP data of 2,325,889 loci across all individuals. Marker/trait associations were developed, and these were further refined using two other populations. Genotyping data from a total of 66,429 individuals were used to discover four highly significant loci affecting this debilitating disease. Using this same approach, worldwide consortiums of bean geneticists could be formed. Many of the same traits are of relevance throughout the world, so the phenotypic data can be shared by the different projects. Since the SNP data can be collated using imputation analysis, it is not required that the same genotyping platform be used. GBS data can be collected from all of the trials, and large genotypic datasets can be generated. Obviously, a single crop such as beans does not have the resources to work with a population of the same size as described above. But the fact that the extent of linkage disequilibrium is much longer in bean than humans, it may not be necessary to have such large populations. From such large trials significant marker/trait associations can be generated that will have relevance to many of the bean improvement programs throughout the world.

The third example is from the newly emerging field of population genomics. The goal of these experiments is to utilize resequencing data to discover causative

loci and use that data to develop new markers for breeding. Again, researchers in rice and maize (*Zea mays* L.) are providing the earliest examples. The first step of a population genomics experiment is to identify a population well suited to the discovery of selected loci of interest. Huang et al. (2012) and Xu et al. (2012) were each searching for loci associated with the domestication of rice. Their populations consisted of both wild, landrace, and cultivated varieties. Within the wild, they evaluated both annual and perennial types, whereas the landrace and cultivars represented the Japonica, Indica, and Aus types. Hufford et al. (2012) were searching for loci associated with corn domestication and breeding improvement and their population consisted of wild, landrace, and improved genotypes. Jiao et al. (2012) limited their analyses to the effects of selection in modern maize, and their population contained a range of inbreds representing breeding efforts from the past several decades.

Capturing genotype data for members of the population is the next step. SNP variation collected from genome sequence data is the preferred genotyping approach. There are two strategies to collecting the data. For well-funded projects, each member of the population is sequenced. Alternatively, DNA pools can be sequenced. These pools are created by mixing DNA samples from individuals within a population. An early example is from chicken, where researchers were looking for signatures of selection associated with the breeding of broilers and layers. Broiler and layer DNA pools were created, and the pools sequenced individually (Rubin et al. 2010). This approach was also recently used to discover loci associated with dog (Axelsson et al. 2013) and pig domestication (Rubin et al. 2012).

The third step in population genomics analysis is to search for regions or loci that are outliers. Often sliding windows are used as the test unit to discover the outliers. The test statistic for outliers varies depending upon the analysis; there is no one accepted standard. There are several test statistics that have been used and each looks for a reduction in diversity or increased differentiation relative to a presumed ancestral population. For the rice studies mentioned above (Huang et al. 2012; Xu et al. 2012), π, a measure of within population nucleotide diversity, was calculated for each window in the wild and landrace populations. Then the ratio $\pi_{wild}/\pi_{landrace}$ was calculated. If a landrace genomic window underwent selection relative to the wild population, this ratio would be large. F_{ST}, a measure of population differentiation introduced by Wright (1951), is another test statistic. A large F_{ST} value indicates great differentiation at that window, with the possibility that a gene that underwent selection is located in the interval.

The final and most challenging step of population genomics is to confirm that the region defined as under selection has a function. To date, this has only been accomplished by comparing the population genomic results with the results of cloned domestication genes. In both the rice (Huang et al. 2012; Xu et al. 2012) and maize (Hufford et al. 2012) studies, previously cloned domestication genes were observed to fall into regions determined to be under selection. These observations validated the approach. Beyond that, these studies tabulated the genes found in selected regions and considered them as candidates for either domestication, and in the case of maize, improvement (Hufford et al. 2012). The number of candidate genes for Indica and Japonica rice was 750 and 439, respectively ($\alpha = 0.025$) (Xu et al. 2012),

while Hufford et al. (2012) defined 468 domestication and 571 improvement genes in corn ($\alpha = 0.10$).

So how these approaches could be used in the improvement of common bean? The major market classes of beans have distinct features. The signatures of these features are buried in the genome sequences, and a comparison of the sequences of representatives of different market classes will identify those differences. These studies can be compared with the mapping data from the large trials described above to correlate the variable regions identified from the population genomics studies with association mapping marker results. If they are correlated, it is highly likely that genome involved in the trait of interest is located in that region. And as always, the genomic sequence data can be used to develop new functional markers for breeding programs.

Phenotyping

In the same way that reliable genotyping is important, good and dependable phenotyping is needed as well in order to have accuracy and consequently, good quality results. Great improvements can be shown in terms of genotyping, especially in terms of efficiency (less cost and time, more samples). In contrast, the way many phenotypic traits are measured continues to be the same in most cases. Consequently, it is faster today to obtain genotypic data points, but cumbersome to obtain phenotypic data at the same pace. The need for high throughput phenotyping is crucial to augment the advancement and successful applications of genomic tools into breeding programs. This is especially important for traits that are difficult to measure such as seed yield, drought tolerance and/or water use efficiency, plant and root architecture, and many other traits measured throughout the phenologic cycle (von Mogel 2013). Screening for disease resistance can be done in the greenhouse for numerous genotypes and then correlate to smaller field evaluations as a way to obtain reliable data. The problem with this approach is that in many cases, the reactions observed in the greenhouse differ from the ones observed in the field because different mechanisms may take place (Myers et al. 1999; Kim et al. 2000).

In the case of abiotic stresses such as drought or soil constraints, the problem is even more complex because of the genetic nature of the target traits, which makes difficult to find direct ways to measure them (Ishitani et al. 2004). In those cases, scientists need to use indirect traits as informative proxies of the traits of interest. In the case of drought, Beebe (2012) provides an extensive review about efforts, methodologies and target traits to phenotype. The authors suggest a 2-way strategy to measure both target shoot traits and target root traits. Some of these traits can be measured in greenhouse experiments while others need to be in the field in order to obtain realistic data that can be exploited in breeding programs. While initial QTL studies have been promising, these have mostly been in a limited number of RIL populations, all so far created from crosses within the Mesoamerican gene pool (Beebe 2012; Schneider et al. 1997). Further studies with populations developed

from crosses between gene pools or from crosses within the Andean gene pool are needed to explore additional diversity for drought resistance QTL alleles, and to analyze the effect of genetic backgrounds on the QTL alleles that have already been identified. In the same way, there is a need for a larger number of polymorphic markers to analyze populations derived within a gene pool. To do this effectively, larger populations are needed for genetic analysis, since most RIL populations in common beans have only been developed with around 100 lines. However, development and maintenance of RILs with large population sizes (e.g. around 300 lines) is more difficult in beans that in other crops such as maize and cereals. If MAS for drought tolerance is to be successful, then understanding the interaction of QTL alleles with multiple genetic backgrounds is important, since breeding programs usually deal with a range of commercial classes and seed colors representing different genetic backgrounds, gene pools, and races.

An additional challenge to the genetic understanding of economically important QTLs is their underlying genetic and mechanistic factors and the interactions among them. These can be regulatory genes such as those governing transcription factors, or structural genes such as those involved in hormone pathways, carbon or nitrogen metabolism, secondary metabolite production, among others. Several non-invasive phenotyping sensors and protocols can be adapted to bean populations and breeding material. Quantifying root growth in relation to resource availability may help to understand how beans exploit below-ground nutrients and interact with their neighbours. Imaging spectroscopy will further our understanding of how leaves and shoots are affected by environmental constraints and of spatial and temporal stress responses of plants and canopies. Non-invasive and spatially resolved measurements of transport processes and resource allocation may be used to better understand and potentially guide transport between plant organs and pods. Finally, single high-throughput phenotyping pipelines are useful, but only the combination of several phenotyping protocols measuring traits at different scales (from semi-controlled environments to field) will significantly contribute to a deeper understanding of the dynamic processes in both individual plants and canopies and eventually, seed yield.

Conclusion

Common bean has a rich history of developments and discoveries in the genomics area and numerous examples of real-life applications of these tools to breeding programs worldwide. Bean breeders and geneticists have been proactive in trying to implement many of these tools into the breeding pipeline with the goal of gaining efficiency during the selection process. Several examples have been exposed in this chapter that illustrates a gain in resources (time, space, and/or costs). The genomic architecture of common bean allows for straightforward studies and interesting applications will come in the near future now that the genome sequence is available. In addition, the new high-throughput genotyping systems now available for common beans will allow the evaluation of thousands of genotypes with high resolution

and accuracy if good phenotypic data is available. We hope that these new tools will help breeders and geneticist to elucidate and have a better understanding of complex traits controlled by multiple genes and how they can be efficiently incorporated into the breeding pipeline.

References

Acosta-Gallegos JA, Kelly JD, Gepts P (2007) Prebreeding in common bean and use of genetic diversity from wild germplasm. Crop Sci 47:S44–S59

Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles ER, Quian Q, Kitano H, Matsuoka M (2005) Cytokinin oxidase regulates rice grain production. Science 309:741–745

Axelsson E, Ratnakumar A, Arendt M-L, Maqbool K, Webster MT, Perloski M, Liberg O, Arnemo JM, Hedhammar A, Lindblad-Toh K (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature 495:360–364

Balardin RS, Kelly JD (1998) Interaction among races of *Colletotrichum lindemuthianum* and diversity in *Phaseolus vulgaris*. J Am Soc Hortic Sci 123:1038–1047

Balding DJ (2006) A tutorial on statistical methods for population association studies. Nat Rev Genet 7:781–791

Barbosa AEAD, Albuquerque ÉVS, Silva MCM, Valencia A, Rocha TL, Grossi-de-Sa MF (2010) Alpha-Amylase inhibitor-1 gene from *Phaseolus vulgaris* expressed in *Coffea arabica* plants inhibits alpha-amylases from the coffee berry borer pest. BMC Biotechnol 10:44

Beaver JS, Osorno JM (2009) Achievements and limitations of contemporary common bean breeding using conventional and molecular approaches. Euphytica 168:145–176

Beebe S (2012) Common bean breeding in the tropics. In: Janick J (ed) Plant breeding reviews 36. Wiley, Hoboken, NJ, pp 357–426

Bennett MD, Leitch IJ (2005) Nuclear DNA amounts in angiosperms: progress, problems and prospects. Ann Bot 95:45–90

Bitocchi E, Nanni L, Bellucci E, Rossi M, Giardini A, Zeuli PS, Logozzo G, Stougaard J, McClean P, Attene G, Papa R (2012) Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. Proc Natl Acad Sci U S A 109:E788–E796

Blair M, Pedraza F, Buendia H, Gaitan E, Beebe S, Gepts P, Thome J (2003) Development of a genome wide anchored microsatellite for common bean (*Phaseolus vulgaris* L). Theor Appl Genet 107:1362–1374

Blair MW, Rodriguez LM, Pedraza F, Morales F, Beebe S (2007) Genetic mapping of the bean olden mosaic geminivirus resistant gene Bgm-1 and linkage with potyvirus resistance in common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 107:1362–1374

Blair MW, Díaz LM, Buendía HF, Duque MC (2009) Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris* L.). Theor Appl Genet 19:955–972

Blair MW, Hurtado N, Chavarro CM, Muñoz-Torres MC, Giraldo MC, Pedraza F, Tomkins J, Wing R (2011) Gene-based SSR markers for common bean (*Phaseolus vulgaris* L.) derived from root and leaf tissue ESTs: an integration of the BMc series. BMC Plant Biol 11:50

Bonfim K, Faria JC, Nogueira EOPL, Mendes ÉA, Francisco JL, Aragão (2009) RNAi-mediated resistance to *Bean golden mosaic virus* in genetically engineered common bean (*Phaseolus vulgaris*). Mol Plant Microbe Interact 20:717–726

Broughton WJ, Hernandez G, Blair M, Beebe S, Gepts P, Vanderleyden J (2003) Beans (*Phaseolus* spp.) – model food legumes. Plant and Soil 252:55–128

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK,

Li H, Mitchell SE, Pressoir G, Peiffer JA, Oropeza-Rosas M, Rocheford TR, Romay MC, Romero S, Salvo S, Sanchez-Villeda H, Sofia da Silva H, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD (2009) The genetic architecture of maize flowering time. Science 325:714–718

Caldwell KS, Russell J, Langridge P, Powell W (2006) Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. Genetics 172:557–567

Câmara CR, Urrea CA, Schlegel V (2013) Pinto Beans (*Phaseolus vulgaris* L.) as a functional food: implications on human health. Agriculture 3:90–111

Cavanagh C, Morell M, Mackay I, Powell W (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Curr Opin Plant Biol 11:215–221

Chen Y (2013) Dissection of agronomic traits in crops by association mapping. In: Lübberstedt T, Varshney RK (eds) Diagnostics in plant breeding. Springer, Dordrecht, pp 119–142

Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos T Roy Soc B 363:557–572

Cook DE, Lee TG, Guo X, Melito S, Wang K, Bayless AM, Wang J, Hughes TJ, Willis DK, Clemente TE, Biers BW, Jiang J, Hudson ME, Bent AF (2012) Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. Science 338:1206–2909

Córdoba J, Chavarro C, Schlueter J, Jackson S, Blair M (2010) Integration of physical and genetic maps of common bean through BAC-derived microsatellite markers. BMC Genomics 11:436

Duvick DN, Cassman KG (1999) Post-green revolution trends in yield potential of temperate maize in the north-central United States. Crop Sci 39:1622–1630

Eapen S (2008) Advances in development of transgenic pulse crops. Biotechnol Adv 26:162–168

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379

Ender M, Terpstra K, Kelly JD (2008) Marker-assisted selection for white mold resistance in common bean. Mol Breed 21:149–157

Evans LT, Fisher RA (1999) Yield Potential: its definition, measurement, and significance. Crop Sci 39:1544–1551

Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, Li X, Zhang Q (2006) GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor Appl Genet 112:1164–1171

Felicetti E, Song Q, Jia G, Cregan P, Bett K, Miklas PN (2012) Simple sequence repeats linked with slow darkening trait in pinto bean discovered by single nucleotide polymorphism assay and whole genome sequencing. Crop Sci 52:1600–1608

Freyre R, Skroch P, Geffroy V, Adam-Blondon A-F, Shirmohamadali A, Johnson W, Llaca V, Nodari R, Pereira P, Tsai S-M, Tohme J, Dron M, Nienhuis J, Vallejos C, Gepts P (1998) Towards an integrated linkage map of common bean. 4. Development of a core map and alignment of RFLP maps. Theor Appl Genet 97:847–856

Freytag GF, Debouck DG (2002) Taxonomy, distribution and ecology of the genus *Phaseolus* (Leguminosae-Papilionoideae) in North America, Mexico and Central America. Sida botanical miscellany, vol 23. Botanical Research Institute of Texas, Fort Worth, TX

Galeano CH, Fernandez AC, Franco-Herrera N, Cichy KA, McClean PE, Vanderleyden J, Blair MW (2011) Saturation of an intra-gene pool linkage map: towards a unified consensus linkage map for fine mapping and synteny analysis in common bean. PLoS One 6:12

Gepts P, Beavis WD, Brummer EC, Shoemaker RC, Stalker HT, Weeden NF, Young ND (2005) Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through genomics conference. Plant Physiol 137:1228–1235

Gepts P, Aragão FJL, de Barros E, Blair MW, Brondani R, Broughton W, Galasso I, Hernández G, Kami J, Lariguet P, McClean P, Melotto M, Miklas P, Pauls P, Pedrosa-Harand A, Timothy P, Sánchez F, Sparvoli F, Yu K (2008) Genomics of *Phaseolus* beans, a major source of dietary protein and micronutrients in the tropics. In: Moore PH, Ming R (eds) Genomics of tropical crop plants. Springer science + Business media, New York, NY, pp 113–143

Gonçalves-Vidigal MC, Cruz AS, Garcia A, Kami J, Vidigal Filho PS, Sousa LL, McClean P, Gepts P, Pastor-Corrales MA (2011) Linkage mapping of the *Phg-1* and *Co-1*^{4} genes for resistance to angular leaf spot and anthracnose in the common bean cultivar AND 277. Theor Appl Genet 122:893–903

Guzmán P, Gilbertson RL, Nodari R, Johnson WC, Temple SR, Mandala D, Gepts P (1995) Characterization of variability in the fungus *Phaeoisariopsis griseola* suggests coevolution with the common bean (*Phaseolus vulgaris* L.). Phytopathology 85:600–607

Hamblin MT, Salas-Fernandez MG, Casa AM, Mitchell SE, Paterson AH, Kresovich S (2005) Equilibrium processes cannot explain high levels of short-85and medium-range linkage disequilibrium in the domesticated grass sorghum bicolor. Genetics 171:1247–1256

Hougaard BK, Madsen LH, Sandal N, Moretzsohn MC, Fredslund J, Schauser L, Nielsen AM, Rohde T, Sato S, Tabata S, Bertioli DJ, Stougaard J (2008) Legume anchor markers link syntenic regions between *Phaseolus vulgaris*, *Lotus japonicus*, *Medicago truncatula* and *Arachis*. Genetics 119:2299–2312

Huang X, Kurata N, Wei X, Wang ZX, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W, Guo Y, Lu Y, Zhou C, Fan D, Weng Q, Zhu C, Huang T, Zhang T, Wang Y, Feng L, Furuumi H, Kubo T, Miyabayashi T, Yuan X, Xu Q, Dong G, Zhan Q, Li C, Fujiyama A, Toyoda A, Lu T, Feng Q, Qian Q, Li J, Han B (2012) A map of rice genome variation reveals the origin of cultivated rice. Nature 490:497–501

Huang BE, Clifford D, Cavanagh C (2013) Selecting subsets of genotyped experimental populations for phenotyping to maximize genetic diversity. Theor Appl Genet 126:379–388

Hufford MB, Xu X, van Heerwaarden J, Pyhajarvi J, Chia J-M, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, Lai J, Morrell PL, Shannon LM, Song C, Springer NM, Swanson-Wagner RA, Tiffin P, Wang J, Zhang G, Doebley J, McMullen MD, Ware D, Buckler ES, Yang S, Ross-Ibarra J (2012) Comparative population genomics of maize domestication and improvement. Nat Genet 44:808–811

Hyten DL, Chol IY, Song QJ, Shoemaker RC, Nelson RL, Costa JM, Specht JE, Cregan PB (2007) Highly variable patterns of linkage disequilibrium in multiple soybean populations. Genetics 175:1937–1944

Hyten DL, Song O, Fickus EW, Quigley CV, Lim JS, Choi IY, Hwang EY, Pastor-Corrales M, Cregan PB (2010) High-throughput SNP discovery and assay development in common bean. BMC Genomics 11:475

IRGSP (International Rice Genome Sequencing Project) (2005) The map-based sequence of the rice genome. Nature 436:793–800

Ishitani M, Raoa I, Wenzlb P, Beebea S, Tohme J (2004) Integration of genomics approach with traditional breeding towards improving abiotic stress adaptation: drought and aluminum toxicity as case studies. Field Crops Res 90:35–45

Jiao Y, Zhao YH, Ren L, Song W, Zeng B, Guo J, Wang B, Liu Z, Chen J, Li W, Zhang M, Xie S, Lai J (2012) Genome-wide genetic changes during modern breeding of maize. Nat Genet 44:812–815

Kelly JD, Vallejo VA (2004) A comprehensive review of the major genes conditioning resistance to anthracnose in common bean. HortScience 39:1196–1207

Kim HS, Hartman GL, Manandhar JB, Graef GL, Steadman JR, Diers BW (2000) Reaction of soybean cultivars to Sclerotinia stem rot in field, greenhouse, and laboratory evaluations. Crop Sci 40:665–669

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R (2009) A Multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. PLoS Genet 5:e1000551. doi:10.1371/journal.pgen.1000551

Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. Trends Plant Sci. doi:10.1016/j.tplants.2006.12.001

Mamidi S, Chikara S, Goos RJ, Hyten DL, Annam D, Moghaddam SM, Lee RK, Cregan PB, McClean PE (2011) Genome-wide association analysis identifies candidate genes associated with iron deficiency chlorosis in soybean. Plant Genome 4:154–164

McClean PE, Lee RK, Otto C, Gepts P, Bassett MJ (2002) Molecular and phenotypic mapping of genes controlling seed coat pattern and color in common bean *(Phaseolus vulgaris* L.). J Heredity 93:148–152

McClean P, Gepts P, Jackson S, Rokshar D, Vance C (2008) Towards a whole genome sequence of common bean (*Phaseolus vulgaris*): background, approaches, applications. Version 1. http://bic.css.msu.edu/_pdf/Bean_Genomics_Status_2008.pdf. Accessed 28 May 2013

McClean PE, Mamidi S, McConnell M, Chikara S, Lee R (2010) Synteny mapping between common bean and soybean reveals extensive blocks of shared loci. BMC Genomics 11:184

McConnell M, Mamidi S, Lee R, Chikara S, Rossi M, Papa R, McClean P (2010) Syntenic relationships among legumes revealed using a gene-based genetic linkage map of common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 121:1103–1116

Miklas PN, Singh SP (2007) Common bean. In: Cole C (ed) Genome mapping and molecular breeding in plants, volume 3. Pulses, sugar and tuber crops. Springer, Heidelberg, pp 1–31

Miklas PN, Stone V, Daly MJ, Stavely JR, Steadman JR, Bassett MJ, Delorme R, Beaver JS (2000) Bacterial, fungal, and viral disease resistance loci mapped in a recombinant inbred common bean population ('Dorado'/XAN 176). J Am Soc Hortic Sci 125:476–481

Miklas PN, Kelly JD, Beebe SE, Blair MW (2006) Common bean breeding for resistance against biotic and abiotic stresses: From classical to MAS breeding. Euphytica 147:105–131

Mkandawire AB, Mabagala RB, Guzmán P, Gepts P, Gilbertson RL (2004) Genetic diversity and pathogenic variation of common blight bacteria (*Xanthomonas campestris* pv. phaseoli and *X. campestris* pv. phaseoli var. fuscans) suggests pathogen coevolution with the common bean. Phytopathology 94:593–603

Moghaddam SM, Song Q, Mamidi S, Lee R, Cregan P, Osorno JM, McClean PE (2013) Developing market class specific indel markers from next generation sequence data in *Phaseolus vulgaris*. Frontiers Plant Sci doi:10.3389/fpls.2013.00251

Morrell PL, Buckler ES, Ross-Ibarra J (2011) Crop genomics: advances and applications. Nat Rev Genet 13:85–96

Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE, Harriman J, Glaubitz JC, Buckler ES, Kresovich S (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc Natl Acad Sci U S A 110:453–458

Myers JR, Gilmore B, Kean D (1999) Correlation between the field and straw test for white mold resistance in common bean. Annu Rep Bean Improv Coop 42:57–58

Naj AC, Jun G, Beecham GW, Wang L-S, Vardarajan BN, Buros J, Gallins PJ, Buxbaum JD, Jarvik GP, Crane PK, Larson EB, Bird TD, Boeve BF, Graff-Radford NR, De Jager PL, Evans D, Schneider JA, Carrasquillo MM, Ertekin-Taner N, Younkin SG, Cruchaga C, Kauwe JSK, Nowotny P, Kramer P, Hardy J, Huentelman MJ, Myers AJ, Barmada MM, Demirci FY, Baldwin CT, Green RC, Rogaeva E, St George-Hyslop P, Arnold SE, Barber R, Beach T, Bigio EH, Bowen JD, Boxer A, Burke JR, Cairns NJ, Carlson CS, Carney RM, Carroll SL, Chui HC, Clark DG, Corneveaux J, Cotman CW, Cummings JL, DeCarli C, DeKosky ST, Diaz-Arrastia R, Dick M, Dickson DW, Ellis WG, Faber KM, Fallon KB, Farlow MR, Ferris S, Frosch MP, Galasko DR, Ganguli M, Gearing M, Geschwind DH, Ghetti B, Gilbert JR, Gilman S, Giordani B, Glass JD, Growdon JH, Hamilton RL, Harrell LE, Head E, Honig LS, Hulette CM, Hyman BT, Jicha GA, Jin L-W, Johnson N, Karlawish J, Karydas A, Kaye JA, Kim R, Koo EH, Kowall NW, Lah JJ, Levey AI, Lieberman AP, Lopez OL, Mack WJ, Marson DC, Martiniuk F, Mash DC, Masliah E, McCormick WC, McCurry SM, McDavid AN, McKee AC, Mesulam M, Miller BL, Miller CA, Miller JW, Parisi JE, Perl DP, Peskind E, Petersen RC, Poon WW, Quinn JF, Rajbhandary RA, Raskind M, Reisberg B, Ringman JM, Roberson ED, Rosenberg RN, Sano M, Schneider LS, Seeley W, Shelanski ML, Slifer MA, Smith CD, Sonnen JA, Spina S, Stern RA, Tanzi RE, Trojanowski JQ, Troncoso JC, Van Deerlin VM, Vinters HV, Vonsattel JP, Weintraub S, Welsh-Bohmer KA, Williamson J, Woltjer RL, Cantwell LB, Dombroski BA, Beekly D, Lunetta KL, Martin ER, Kamboh MI, Saykin AJ, Reiman EM, Bennett DA, Morris JC, Montine TJ, Goate AM, Blacker D, Tsuang DW, Hakonarson H, Kukull WA, Foroud TM, Haines JL, Mayeux R, Pericak-Vance MA, Farrer LA, Schellenberg GD (2011) Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset alzheimer's disease. Nat Genet 43:436–441

Papa R, Bellucci E, Rossi M, Leonardi S, Rau D, Gepts P, Nanni L, Attene G (2007) Tagging the signatures of domestication in common bean (*Phaseolus vulgaris)* by means of pooled DNA samples. Ann Bot 100:1039–1051

Rafalski JA (2010) Association genetics in crop improvement. Curr Opin Plant Biol 13:174–180

Ramirez M, Graham MA, Blanco-Lopez L, Silvente S, Medrano-Soto A, Blair MW, Hernandez G, Vance CP, Lara M (2005) Sequencing and analysis of common bean ESTs. Building a foundation for functional genomics. Plant Physiol 137:1211–1227

Ray DK, Ramankutty N, Mueller ND, West PC, Foley JA (2012) Recent patterns of crop yield growth and stagnation. Nat Commun 3:1293. doi:10.1038/ncomms2296

Repinski SL, Kwak M, Gepts P (2012) The common bean growth habit gene *PvTFL1y* is a functional homolog of Arabidopsis *TFL1*. Theor Appl Genet 124:1539–1547

Rivkin MI, Vallejos CE, McClean PE (1999) Disease-resistance related sequences in common bean. Genome 42:41–47

Rubin C, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S, Hallböök F, Besnier F, Carlborg O, Bed'hom B, Tixier-Boichard M, Jensen P, Siegel P, Lindblad-Toh K, Andersson L (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464:587–591

Rubin CJ, Megens HJ, Martinez Barrio A, Gaqbool K, Sayyab S, Schwochow D, Wang C, Carlborg O, Jern P, Jorgensen CB, Archibald AL, Fredholm M, Groenen MA, Andersson L (2012) Strong signatures of selection in the domestic pig genome. Proc Natl Acad Sci U S A 109:19529–19536

Sandlin CM, Steadman JR, Araya CM, Coyne DP (1999) Isolates of *Uromyces appendiculatus* with specific virulence to landraces of *Phaseolus vulgaris* of Andean origin. Plant Dis 83:108–113

Schlueter JA, Goicoechea JL, Collura K, Gill N, Lin J, Yu Y, Kudrna D, Zuccolo A, Vallejos CE, Muñoz-Torres M, Blair MW, Tohme J, Tomkins J, McClean P, Wing RA, Jackson SA (2008) BAC-end Sequence analysis and a draft physical map of the common bean (*Phaseolus vulgaris* L.) genome. Trop Plant Biol 1:40–48

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the paleopolyploid soybean. Nature 463:178–183

Schneider K, Brothers M, Kelly JD (1997) Marker-assisted selection to improve drought resistance in common bean. Crop Sci 37:51–60

Shannon S, Meeks-Wagner DR (1991) A mutation in the Arabidopsis *TFL1* gene affects inflorescence meristem development. Plant Cell 3:877–892

Shi C, Navabi A, Yu K (2011) Association mapping of common bacterial blight resistance QTL in Ontario bean breeding populations. BMC Plant Biol 11:52

Song XJ, Huang W, Shi M, Zhu MZ, Lin HX (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nat Genet 39:623–630

Soule M, Porter L, Medina J, Santana GP, Blair MW, Miklas PN (2011) Comparative QTL map for white mold resistance in common bean, and characterization of partial resistance in dry bean lines VA19 and I9365-3. Crop Sci 51:123–139

Singh SP, Gepts P, Debouck DG (1991). Races of common bean (Phaseolus vulgaris, Fabaceae). Econ Bot 45:379–396

Tian Z, Wang X, Lee R, Li Y, Specht JE, Nelson RL, McClean PE, Qiu L, Ma J (2010) Artificial selection for determinate growth habit in soybean. Proc Natl Acad Sci U S A 19:8563–8568

USDA-ERS (2013) Vegetables and melons outlook—dry edible beans. United States Department of Agriculture, Economic Research Service, Washington, DC

Vallad G, Rivkin M, Vallejos CE, McClean PE (2001) Cloning and homology modeling of a Pto-like protein kinase family of common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 103:1046–1058

Vallejos CE, Sakiyama NS, Chase CD (1992) A molecular marker-based linkage map of *Phaseolus vulgaris* L. Genetics 131:733–740

Vandemark GJ, Brick MA, Kelly JD, Osorno JM, Urrea CA (2013) Yield gains in edible grain legumes. In: Smith S (ed) Yield gains in major U.S. field crops. Madison WI, Crop Science Soc. of America (CSSA)

Von Mogel KH (2013) Taking the phenomics revolution into the field. CSA News Mag 58:4–10

Wang C, Chen S, Yu S (2011) Functional markers from multiple loci in *GS3* for fine marker-assisted selection of grain length in rice. Theor Appl Genet 122:905–913

Wright S (1951) The genetical structure of populations. Ann Eugen 15:322–354

Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L, Li J, He W, Zhang G, Zheng X, Zhang F, Li Y, Yu C, Kristiansen K, Zhang X, Wang J, Wright M, McCouch S, Nielsen R, Wang J, Wang W (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotechnol 30:105–111

Yamanaka S, Nakamura I, Watanabe KN, Sato YI (2004) Identification of SNPs in the waxy gene among glutinous rice cultivars and their evolutionary significance during the domestication process of rice. Theor Appl Genet 108:1200–1204

Yang SS, Valdés-López O, Xu WW, Bucciarelli B, Gronwald JW, Hernández G, Vance CP (2010) Transcript profiling of common bean (*Phaseolus vulgaris* L.) using the GeneChip® soybean genome array: optimizing analysis by masking biased probes. BMC Plant Biol 10:85

Yang X, Gao S, Xu S, Zhang Z, Prasanna BM, Li L, Li J, Yan J (2011) Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. Mol Breeding 28:511–526

Yu K, Park SJ, Poysa V (2000) Marker-assisted selection of common beans for resistance to common bacterial blight: efficacy and economics. Plant Breed 119:411–415

Yuste-Lisbona FJ, Santalla M, Capel C, Garcia-Alcazar, De La Fuente M, Capel C, De Ron A, Lozano R (2012) Marker-based linkage map of Andean common bean (*Phaseolus vulgaris* L) and mapping of QTLs underlying popping ability traits. BMC Plant Biol 12:136

Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, McCouch SR (2011) Genome-wide association mapping reveals rich genetic architecture of complex traits in *Oryza sativa*. Nat Commun 2:467

Zhu H, Choi H, Cook DR, Shoemaker RC (2005) Bridging model and crop legumes through comparative genomics. Plant Physiol 137:1189–1196

Chapter 10
Pulses Biofortification in Genomic Era: Multidisciplinary Opportunities and Challenges

Dil Thavarajah, Pushparajah Thavarajah, and Debjyoti Sen Gupta

Abstract Agricultural production systems driven by green revolutionary efforts have resulted in the displacement of traditional food crops that provided greater levels of protein and essential micronutrients. This has led to half of the world's population being deficient in essential micronutrients, with millions lacking in daily protein and micronutrient intakes. Biofortification of many commonly eaten staple foods is viewed as a sustainable solution to combat global micronutrient malnutrition. However, biofortification efforts with pulse crops [mainly lentils (*Lens culinaris* L.), field pea (*Pisum sativum* L.), and chickpea (*Cicer arietinum* L.)] have been limited to certain regions in North America and no global pulse biofortification initiative exists. The majority of the world's population lives in Asia and Africa, where there is an urgent need to produce micronutrient and protein rich pulses to prevent micronutrient malnutrition deficiencies. This chapter reviews the last 10 years of literature on pulse genetic biofortification, discusses current pulse biofortification research efforts in the USA and other countries, and suggests urgent pulse biofortification efforts involving modern genomic tools and techniques along with the sound phenotyping targeted at finding sustainable solutions for regions with the greatest micronutrient deficiencies.

Keywords Biofortification • Micronutrient malnutrition • Pulse crops • Food systems • Iron • Zinc • Selenium • Beta-carotene • Molecular mapping

D. Thavarajah, Ph.D. (✉) • P. Thavarajah, Ph.D.
School of Food Systems, North Dakota State University, Dept. 7640,
NDSU 1250 Bolley Drive, 208 Harris Hall, Fargo, ND 58108, USA
e-mail: dilrukhshi.thavarajah@ndsu.edu

D.S. Gupta, M.Sc. (Ag.)
Cereal Science Graduate Program, Department of Plant Sciences,
North Dakota State University, Fargo, ND, USA

S. Gupta et al. (eds.), *Legumes in the Omic Era*, DOI 10.1007/978-1-4614-8370-0_10,

Introduction

World agriculture aims to provide sufficient food energy and nutrients for the general well-being and health of human populations. Over time, agricultural development has been one of the greatest achievements of mankind. One key example, the green revolution, was a science-driven agricultural endeavor that aimed to provide adequate food energy for the world's populations, including more calories per person. Many countries around the world greatly benefited from this agricultural revolution that expanded global production of staple food crops, especially cereals. The green revolution efforts that began in the early 1960s achieved adequate world food production or calories/protein per person in less than two decades. However, the achievements with fast growing, high yielding cereals displaced traditional food crops that provided greater levels of essential micronutrients. In addition, this trend to replace traditional food crops failed to link agricultural production to human nutrition and health, and has led to the development of food systems that do not meet human nutritional needs or are simply unhealthy. As a result, in recent decades the global prevalence of iron (Fe) deficiency anemia has increased to 40 % among women and children. In India, the prevalence of Fe deficiency anemia increased to 70 % during the green revolution period (Welch and Graham 1999).

Micronutrient malnutrition, also known as "hidden hunger", affects more than two billion people worldwide. Particularly vulnerable are women and preschool children in South Asia. Estimates indicate that over 60 % of the world's seven billion people are Fe deficient, over 30 % are zinc (Zn) deficient, 30 % are iodine (I) deficient, and more than 15 % are selenium (Se) deficient. Beta-carotene and folate deficiencies are also increasing; approximately three million children around the world develop vitamin A deficiency and every year more than half a million children lose their eyesight. The concept of biofortification (micronutrient enrichment through conventional plant breeding and modern biotechnology) emerged in the late 1990s and is currently considered one of the most sustainable agricultural approaches to combat global micronutrient malnutrition (Welch 2002).

The development of micronutrient-enriched staple food crops may be an effective and sustainable means to increase micronutrient intake to support general human health. Therefore, biofortification of traditional pulses including lentils (*Lens culinaris* L.), field pea (*Pisum sativum* L.), and chickpea (*Cicer arietinum* L.) with highly bioavailable Fe, Zn, Se, I, beta-carotene, and folic acid is urgently needed to address chronic diseases linked to micronutrient malnutrition around the world. Recently, the genetic potential for biofortification of bioavailable Fe, Zn, and provitamin A has been reported for the edible portions of several staple food crops, including rice (*Oryza sativa* L.), wheat (*Triticum* sp.), maize (*Zea mays* L.), common bean (*Phaseolus vulgaris* L.), sweet potato (*Ipomoea batatas* L.), and cassava (*Manihot esculenta* C.) (Welch 2002). This book chapter outlines current North American pulse crop research efforts and future pulse crop "biofortification" needs in the major pulse producing regions in the world in light of recent available phenotyping tools and techniques.

Food Based Solutions: Biofortification

The poor health of billions of people and more than two-thirds of the deaths of young children around the world today are associated with micronutrient malnutrition-related metabolic disorders. Reduced intakes of essential micronutrients below the recommended daily allowance have contributed to increased rates of morbidity and mortality, increased rates of learning disorders, shorter life spans, poorer growth and cognitive development, greater impact of infectious diseases, and the increased prevalence of other physical disabilities in developing nations. In developed countries, micronutrient malnutrition contributes to cardiovascular diseases, cancer, stroke, diabetes, Fe deficiency anemia, low birth weight, and mental retardation. Failure to link agricultural production with human nutrition and health has led to the development of unhealthy food systems that result in micronutrient malnutrition. This is currently leading to both basic malnutrition challenges and chronic diseases linked to high caloric intake; the effects are global but are most evident in South Asia, Africa, and Latin America. Poor diets and nutrition not only cause death around the world but also create huge economic burdens in healthcare with negative consequences on child development and long-term direct impacts on sustainable development. The function of sustainable agriculture systems is to provide food energy and nutrients for humans, thus supporting their health and general well-being and ultimately advancing sustainable development. Resolution of these multi-layered challenges will require the development of a new agriculture paradigm aimed at achieving sustainable food solutions to combat global micronutrient malnutrition through improving food nutritional quality and meet increasing global food demands (Welch and Graham 2004).

Biofortification, or the development of micronutrient-enriched staple food crops through traditional plant breeding methods in conjunction with modern molecular biological techniques, is a powerful and sustainable intervention method to combat global micronutrient deficiencies (Welch 2002). Natural enrichment of traditional food crops with highly bioavailable Fe, Zn, Se, I, beta-carotene, and folic acid is urgently needed to address micronutrient malnutrition and obesity-linked chronic diseases in both developed countries and the developing world. Past attempts to address micronutrient malnutrition have included dietary supplements, food fortification, diet diversification, and supplementation. Unlike supplementation and fortification, which add ongoing costs to consumers, biofortification offers the opportunity to change crop nutritional value within the production system and in ways that have little or no impact on consumer cost. In addition to human nutritional benefits, biofortification can also increase crop production in terms of grain yield and biomass. Most cultivated land areas in the developing world are poor in plant bioavailable essential micronutrients. Studies indicate biofortified crop seeds generally perform well in poorly fertilized soils in developing nations compared to non-biofortified or regular seeds (Welch and Graham 1999; Welch 2002). This is further evidence that biofortified seeds and biofortification efforts may be a sustainable solution to global micronutrient malnutrition.

The micronutrient density of staple food crops can be increased in many ways. Both agronomic and genetic biofortification is effective for enriching micronutrients depending on the target element; however, genetic biofortification is more economical and sustainable in the long term, especially for resource poor countries. Agronomic biofortification is simply a low dosage application of micronutrients to increase the mineral concentration in the edible portion of food crops. Successful case studies have been reported for Se in Finland, New Zealand, Turkey, and South-east Asia (Wang et al. 1995). Agronomic biofortification can be achieved by application of the appropriate mineral forms through foliar or soil application at planting. Agronomic biofortification with Fe, either by addition of soluble ferrous fertilizer to the soil or by foliar application, is not effective for enrichment of Fe due to rapid oxidation and low mobility of Fe in phloem. In such situations, genetic biofortification is an effective solution. However, such efforts will require a better understanding of the global situation, food systems, the complexities of a balanced human diet, and the government policies that support global biofortification activities. Current global biofortification efforts are limited to a few selected staple crops. The biofortification of nitrogen fixing pulses grown as staple crops in Asian countries has not yet been studied, despite the fact that such efforts could address both malnutrition and obesity-linked chronic disease challenges at the same time.

The Physiology of Mineral Uptake in Plants

The physiological role of micronutrient uptake and the efficiency of this process in staple food crops are not yet clearly understood. Generally, mineral (e.g., Fe, Zn) uptake in plants is governed by homoeostatic mechanisms that regulate metal absorption, translocation, and redistribution to provide adequate amounts for completion of the life cycle but prevent accumulation to toxic levels. Several naturally occurring barriers are found in plants to facilitate homoeostatic mechanisms: (a) the root-soil interface (i.e., rhizosphere); (b) transporters and ion channels present in the root-cell plasma membrane; (c) efficient translocation and accumulation in edible plant organs; and (d) bioavailability of minerals. To increase the accumulation of bioavailable forms of micronutrients in seeds, all four of these physiological processes should be clearly understood for each element (Welch and Graham 2004).

Iron uptake in plants is highly regulated in order to supply sufficient amounts for optimal growth and development. Insufficient Fe uptake leads to Fe-deficiency symptoms, including interveinal chlorosis, necrosis in leaves, and finally reduction of biomass and grain yield. Plants regulate Fe uptake to prevent excess accumulation of hydroxyl radicals in the cells. These hydroxyl radicals are formed during the reduction of molecular oxygen and can damage cellular components, such as DNA, proteins, lipids, and sugars. Generally, plants require at least 10^{-9}–10^{-4} mol/l Fe to achieve optimal growth and development; however, maintaining this concentration in plants is very challenging due to the low solubility of Fe in soil solution (Römheld and Schaaf 2004). Plants acquire Fe via two strategies: strategy I (via the Fe^{3+} chelate reductase) for dicots and chelation-based strategy II (via exchange chelation

with phytosiderophores) for monocots such as corn, wheat, and rice. Genes involved in both of these mechanisms have been identified from plants, including rice, maize, and *Arabidopsis* (White and Broadley 2009).

Zinc is another element essential to many plant proteins but also toxic in excess. Plants acquire Zn primarily as a divalent cation from the soil solution but it can also be complexed with organic acids present in the rhizosphere. Plants typically use both symplastic and apoplastic fluxes processes to transport Zn from root to shoot (Broadley et al. 2007).

Selenium (Se) is an essential element for mammals but has not been considered an essential element for higher plants. The biochemistry of Se in plants has been reviewed (Ellis and Salt 2003; Combs 2001). Plants uptake Se from the soil primarily as inorganic Se (selenate or selenite) and translocate it to the chloroplast via the sulfur assimilation pathway. In the chloroplast, adenosine-5-phosphoselenate is formed by the activation of ATP sulfurylase. This selenate is then reduced to selenide, which reacts with serine to form selenocysteines (SeCys) and is further metabolized to other organic Se forms. Selenomethionine (SeMet) and SeCys, a homologue of dimeric cysteine in which two Se atoms replaces the disulfide bond, are the major organic forms of Se found in legumes and cereals (Wu et al. 1997).

Micronutrients and Human Health

Every human requires the following nutrients for normal physiological functions and general health: elements boron, chromium, copper, fluorine, iron, manganese, molybdenum, nickel, selenium, silicon, and zinc (Table 10.1); water soluble vitamins

Table 10.1 Recommended nutrient intake for males and females of 19–50 years

Nutrient	Assessment	Female	Male
Calcium (mg/day)	RDA	1,000	1,000
Chromium (μg/day)	AIs	25	35
Copper (μg/day)	RDA	900	900
Fluoride (mg/day)	Ais	3	4
Iodine (μg/day)	RDA	150	150
Iron (mg/day)	RDA	18	8
Magnesium (mg/day)	RDA	310–320	400–420
Manganese (mg/day)	Ais	1.8	2.3
Molybdenum (mg/day)	RDA	45	45
Phosphorus (mg/day)	RDA	700	700
Selenium (μg/day)	RDA	55	55
Zinc (mg/day)	RDA	8	11
Potassium (g/day)	Ais	4.7	4.7
Sodium (g/day)	Ais	1.5	1.5
Chloride (g/day)	Ais	2.3	2.3

Based on data from Welch and Graham 2004; Dietary Reference Intakes 2012

RDA recommended dietary allowance, *Ais* adequate intakes

Table 10.2 Food matrix factors present in pulse crops that promote or inhibit micronutrient bioavailability

Food matrix factor	Nutrient	Major dietary source
Promoters		
1. Prebiotics: inulin and fructans	Fe, Zn, Ca	Lentils, chicory, garlic
2. Beta-carotene	Fe, Zn	Lentil, pea, chickpea, green and orange vegetables
3. Selenium	I	Lentil, pea, chickpea, sea food
4. Organic acids: ascorbic acid	Fe, Zn	Lentils, fresh fruits and vegetables
5. Amino acids	Fe, Zn	Animal meat
Inhibitors		
1. Phytic acid	Fe, Zn, Ca	All legumes, cereals
2. Fiber	Fe, Zn	All legumes, cereals
3. Haemagglutinins	Fe, Zn	Most legumes, wheat
4. Phenolics	Fe, Zn	All legumes
5. Heavy metals	Zn	Contaminated legumes and leafy vegetables

(Based on data from Welch and Graham 2004; Thavarajah and Thavarajah 2012a, b)

(ascorbic acid, biotin, cobalamin, folic acid, niacin, pantothenic acid, pyroxidine, riboflavin, thiamin); and fat-soluble vitamins (retinoic acid, calciferol, tocopherol, phylloquinone, and menoquinone (Graham and Welch 2000). Plant foods are able to supply adequate amounts of these, with the exception of cobalamin that comes from animal-based diets. The nutritional value of plant-based diets depends on both individual nutrient concentration and bioavailability, the latter of which determines how much is truly absorbed by the human body for its metabolic and physiological functions. Bioavailability of a micronutrient is governed by many factors: host, digestive environments, and the presence of mineral absorption promoters and inhibitors in a food. Absorption promoters, such as ascorbic acid, carotenoids, fructooligosaccharides, certain fibers, sulfur amino acids, and meat factors, increase Fe absorption in the human digestive system; phytic acid and polyphenols in plant-based food are the major inhibitory factors of Fe and Zn bioavailability (Table 10.2).

Recently, the genetic potential of increased concentration and bioavailability of Fe, Zn, and provitamin A (carotenoids) has been studied in rice, wheat, common bean, maize, sweet potato, and cassava (HarvestPlus 2012). Before the establishment of the HarvestPlus research program, Welch and Graham (2004) indicated critical factors that must be considered before releasing new lines of micronutrient enriched staple food crops: (1) the grain yield must be increased or maintained; (2) the potential improvements to human health from due to the enriched crops should be significant; (3) the micronutrient enrichment traits should be stable across different environments or growing locations; (4) the bioavailability of the micronutrients present in the enriched crop lines must be demonstrated using a human model under the normal conditions; and (5) the influence of consumer preference must be addressed to ensure the maximum impact on human health is achieved. In 2012, after several years of research and development, HarvestPlus released several micronutrient enriched lines of rice, wheat, common bean, maize, and sweet potato. However, limited biofortification research efforts have been carried out on pulse crops, including lentil, field pea, and chickpea.

Past and Current Genetic Biofortification Efforts on Pulses

Pulses are a part of the daily diet of many vegetarians as well as people in developing countries. Pulses are rich in protein (20–30 %) and an excellent source of dietary fiber, low molecular weight carbohydrates, essential amino acids, polyunsaturated fatty acids, minerals, and vitamins (Bhatty 1988). Lentil is a traditional pulse crop grown mostly in low-rainfall, dryland cropping systems in rotation with cereals. Currently, lentil is grown in more than 50 countries, with annual world production of approximately 4 M tons. More than 85 % of world lentil production occurs in five specific regions: India, Nepal, and Bangladesh (32 %); western Canada (29 %); Turkey and northern Syria (18 %); Australia (4 %); and the Midwestern region of the USA including North Dakota, South Dakota, and eastern Montana (3 %) (FAOSTAT 2010). Field pea is another important traditional pulse crop mostly grown in cool temperate climates with moist black soils. Annual world pea production is approximately 10–12 M tons, more than 10 % of which is produced in the USA. Chickpea is a third important pulse crop. India produces more than 90 % of the world's chickpeas; only 5 % is grown in the USA and Canada.

Among these three pulse crops, lentil production has seen marked increases of 6.8 % annually, mainly as a result of expanded production in Canada and the USA (FAOSTAT 2010). The Province of Saskatchewan in Canada produces the majority of the world's red lentils. The Midwest region of the USA, including North Dakota, South Dakota, and eastern Montana, is emerging as a major lentil producing region, primarily due to superior yields and an excellent fit in existing crop rotations. Over the last two decades, lentil and pea crops have been significantly increasing in the northern plains region of the US, including North Dakota, South Dakota, and eastern Montana. During this time, the pulse production area has increased from less than 10,000 acres to nearly 900,000 acres (FAOSTAT 2010). Currently, research emphasis on pulse crops grown in North America mainly focuses on improving grain yield, disease tolerance, and nutritional quality, and in depth biofortification research on pulse crops is limited. As of August 2012, approximately 437 research papers were published with the key word "biofortification" in the Science Direct database. Among them, only 32 reported on lentils, 37 on chickpea, and 53 on field pea. This limited literature is evidence that more research is required on pulse biofortification towards finding solutions towards global micronutrient malnutrition.

Within the biofortification framework, lentil breeding programs around the world are working together to increase Fe and Zn concentration and bioavailability to combat global micronutrient malnutrition. As mentioned above, biofortification can improve crop nutritional value with minimal impact on consumer cost. The International Center for Agricultural Research in the Dry Areas (ICARDA) has created a composite collection of more than 1,000 lentil lines to understand the genetic diversity with respect to different nutritional traits, including Fe and Zn accumulation. Sarker et al. (2007) evaluated more than 1,600 lentil genotypes including land races, wild types, and breeding lines of red and green lentils for Fe and Zn biofortification, and reported that lentil lines contained 43–132 mg/kg Fe and 22–78 mg/kg Zn. A later study by Baum et al. (2008) indicates mineral

concentrations in lentil lines ranging from 41–109 mg/kg for Fe and 22–78 mg/kg for Zn. Numerous high Fe and/or Zn lentil genotypes have been released: ILL 5883 (73 mg/kg Fe) and ILL 6994 (72 mg/kg Fe) in Syria; ILL 7711 (74 mg/kg Fe) in Portugal; Alemaya (82 mg/kg Fe, 66 mg/kg Zn) in Ethiopia; Meyveci-2001 (53 mg/kg of Zn) in Turkey; and Sisir (98 mg/kg Fe, 64 mg/kg Zn), Khajurah-2 (94 mg/kg Fe, 54 mg/kg Zn), Barimasur-5 (86 mg/kg Fe, 59 mg/kg Zn), and Barimasur-6 (86 mg/kg Fe, 63 mg/kg Zn) in Nepal.

Research in Canada has highlighted the Fe, Zn, and Se values of pulse crops. More than 900 lentil samples including modern high yielding genotypes grown in Saskatchewan, Canada were evaluated for possible genetic potential for Fe, Zn, and Se biofortification towards improved human health. Considerable genetic variation exists for these minerals, with higher broad sense heritability for Fe and Zn. These lentils contained 73–90 mg/kg of Fe and 44–54 mg/kg of Zn. These results also indicated that Fe and Zn concentrations in lentils are governed by growing location, genotype, year by location, and genotype by location effects. High Fe genotypes CDC Rosetown (90 mg/kg of Fe), CDC Blaze (83 mg/kg of Fe), and CDC Impact (85 mg/kg of Fe) would be potential candidates for future Fe biofortification genetic research aimed at developing suitable Fe enriched lentil lines (Thavarajah et al. 2008, 2009).

Total Se levels in the Canadian lentil samples ranged 425–673 μg/kg, but some genotypes had 40–50 % more Se than others. CDC Robin (extra small red lentil; 672 μg/kg), CDC Sedley (large green; 612 μg/kg), and CDC Grandora (large green; 612 μg/kg) showed higher Se concentrations than the small green lentil genotype Eston (425 μg/kg). Further screening of Se uptake of lentil germplasm by the Canadian lentil program indicated that lentil lines PI330937 and ILL 7537 had higher Se uptake compared to Eston (Thavarajah et al. 2011). Comprehensive data collected by the Canadian lentil breeding program clearly indicates that uptake of Se in lentil seeds is clearly affected by soil and environmental conditions, but that increased lentil Se uptake would be possible through selective breeding.

Limited data are available on the Fe, Zn, and Se content of field pea. Concentrations of 45–49 mg kg^{-1} Fe and 32–35 mg kg^{-1} Zn have been reported for western Canadian-grown field peas (Gawalko et al. 2009). Amarakoon et al. (2012) reported that field peas grown in seven locations in North Dakota, USA, were naturally rich in Fe (46–54 mg kg^{-1}) and Zn (39–63 mg kg^{-1}). Similar results have been reported on field pea grown in Saskatchewan, Canada (Thavarajah et al. 2011). Data from most of these studies were limited to one growing season or a few genotypes. Therefore, future field studies are required to understand the true genetic and genetic × environment interaction effects over several growing seasons, genotypes, and locations. However, these few studies do provide baseline information for biofortification research efforts directed at field pea grown in North America.

Similar to other pulse crops, biofortification research efforts on chickpea have been limited. Ten Canadian grown chickpea cultivars contained 77–112 mg/kg Fe and 29–50 mg/kg Zn (Bueckert et al. 2011). Chickpea grown in Spain had comparably lower values (66 mg/kg of Fe, 35 mg/kg of Zn) (Viadel et al. 2006). However, chickpea grown in the US regional variety trials was found to be a rich

source of Fe (46–67 mg/kg) and particularly Zn (37–74 mg/kg). These results are generally higher than values for samples collected from 2011 US National Pulse Quality Survey (38–47 mg/kg Fe, 20–26 mg/kg Zn) (Thavarajah and Thavarajah 2012a). Overall, the considerably variability underscores the need for more data on chickpea to inform biofortification research.

Food Matrix Factors: Antinutrients

Plant-based foods contain various antinutrients that can reduce the bioavailability of dietary non-heme Fe, Zn, and other nutrients to humans and animals. Some dietary organic acids, amino acids, long chain fatty acids, beta-carotene, and fructooligosaccharides promote the bioavailability of Fe and Zn in the presence of antinutrients, including phytic acid (PA) and polyphenols. The concentration of these promoter and inhibitor compounds in any food crop is influenced by both genetic and environmental factors. Modern plant breeding and molecular biology tools now make it possible to reduce antinutrients, such as phytic acid, or increase the concentrations of promoter substances, such as beta-carotene and ascorbic acid, in plant foods (Table 10.2). Although phytic acid is viewed as an antinutrient, it also has useful roles in plant metabolism and human health, having been associated with reductions in the incidence of chronic heart diseases and cancers. Therefore, efforts to reduce PA concentrations (i.e., low PA mutant development) must be undertaken with caution and in consideration of the optimum balance between seed micronutrient retention and bioavailability.

The bioavailability of minerals present in plant-based diets is greatly affected by the overall composition of the diet. Food processing and preparation techniques can also determine the amount of bioavailable minerals in plant-based diets. Enrichment with prebiotics, beta-carotene, and ascorbic acid has been shown to enhance the bioavailability of non-haem Fe in human plant-based diets (Welch 2002). Prebiotics enhance Fe bioavailability as a result of biological fermentation of short chain polymers (e.g., inulin and fructooligosaccharides) by natural microflora present in the colon. Fermentation may decrease the pH of the luminal content to favor reduction of Fe(III) to Fe(II), stimulate proliferation of epithelial cells, and potentially stimulate expression of mineral-transport protein in human epithelial cells (Yeung et al. 2005). Addition of vitamin A or beta-carotene can improve Fe bioavailability from plant-based foods (e.g., rice, wheat, corn). For example, beta-carotene can overcome the inhibitory effect of phytic acid polyphenols present in a plant-based diet (Garcia-Casal et al. 2000).

Analysis of lentil food matrix components, along with cell culture and preliminary human nutrition studies, reveals clear mineral absorption promoter and inhibitor roles in modulating the levels of mineral bioavailability. Lentils contain high levels of Fe absorption promoters, such as prebiotics and beta-carotene, and are low in antinutrients, such as phytic acid and polyphenols (Thavarajah and Thavarajah 2012b). Molar ratios of PA:Fe above 10 lead to reduced human Fe bioavailability

(Ariza-Nieto et al. 2007). Pulses are naturally low in PA (5.1–7.3 mg/g), and lower than "low phytic acid mutants" developed to date for any other crops (Thavarajah et al. 2009; Thavarajah and Thavarajah 2012b; Amarakoon et al. 2012).

Limitations and Future Scope of Biofortification Using Modern Tools of Genomics

To address micronutrient malnutrition, biofortification of staple foods with highly bioavailable micronutrients is urgently needed. This can be achieved by increasing micronutrient concentrations in edible portions and exploiting bioactive compounds to promote gastrointestinal health toward increased mineral absorption. Our experience with pulses shows that development of nutritionally-superior varieties is a feasible goal. However, the limited amount of pulse biofortification research in major pulse producing and consuming countries may prevent the full micronutrient enrichment potential of pulse crops to be reached. Development of biofortified pulse crops is not only essential for developing whole food-based solutions to micronutrient malnutrition but also necessary for improving agricultural productivities and sustainable development. Nutritionally superior pulses could receive greater consumer acceptance and possibly demand premium pricing in health- and environmentally-conscious global pulse markets.

Biofortification will have a greater impact if the biofortified nutrients are highly bioavailable. The limited pulse biofortification research that has been conducted during the last decade indicates that breeding for micronutrient-rich pulses with high bioavailability may be possible which was predicted by Frossard et al. in 2000. Research efforts aimed at understanding pulse food matrix factors and selection of improved micronutrient bioavailable genetic material could provide the means for developing highly bioavailable and biofortified pulse crops. Highly bioavailable micronutrient pulses may facilitate the delivery of daily nutrient requirements in smaller portion sizes. Highly populated countries with limited pulse production may still achieve micronutrient delivery through improved bioavailability of biofortified pulse crops.

Ghandilyan et al. 2006 suggested forecast of the future of biofortification research efforts as within few recent years the new subject of nutrigenomics came into existence using the potential of genomic tools and techniques for genetic biofortification in crop plants. The use of molecular marker-assisted selection in pulse crop biofortification research efforts has been initiated in parts of the USA, Canada, and CGIAR's around the world. These research initiatives have been presented at international pulse conferences to highlight the promise of this technology. For development of molecular markers linked with the high concentration of micronutrient loci initial large scale evaluation of available germplasm sets of different food legumes is essential as this in turn a prerequisite to develop suitable mapping population (Talukder et al. 2010; Beebe et al. 2000) and for this, state of the art phenotyping facility for micronutrient analysis is required. Presently, Pulse Quality and Nutrition Laboratory at NDSU, Fargo, USA is carrying out such large scale micronutrient analysis including their bioavailability for various pulses including lentils, field

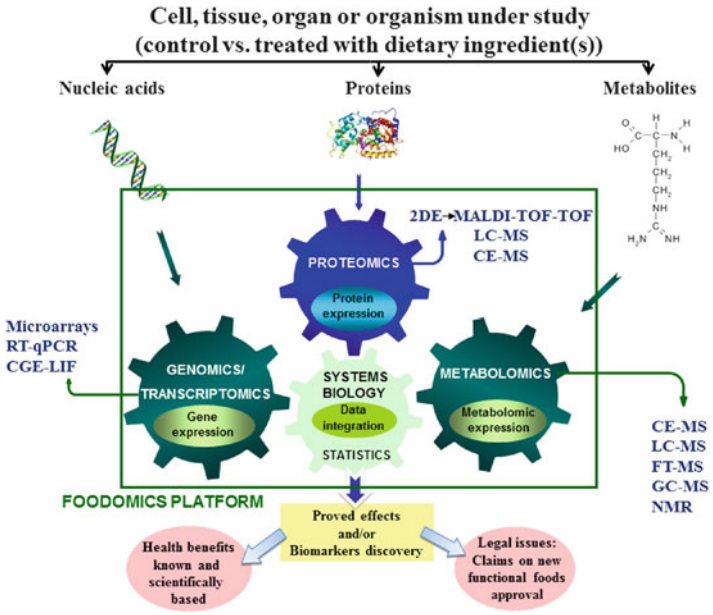


Fig. 10.1 Ideal foodomics platform to study the health benefits from the food constituents on any cell, tissue, organ or organism (Modified from Ibáñez C, Valdés A, García-Cañas V, et al. Global Foodomics strategy to investigate the health benefits of dietary constituents. J. Chromatogr., A 2012; 1248: 139–153. With permission from Elsevier)

peas and chickpeas. In case of few studies so far conducted to map and tag the gene(s)/QTL(s) controlling micronutrient status in legumes and model plants mostly found to be having quantitative mode of inheritance and resulting in identification of gene(s)/QTL(s) capable of explaining moderate amount of phenotypic variation for micronutrient concentration [Sompong et al. 2012 (for phytic acid in mungbean), Blair et al. 2005; Gelin et al. 2007; Cichy et al. 2009; Blair et al. 2010a, b; (for Fe and Zn in commonbean), Sankaran et al. 2009 (for several mineral elements in *Medicago truncatula*), Waters and Grusak 2008 (for several seed mineral contents in *Arabidopsis thaliana*), Walker et al. 2006 (for phytic acid in soybean)]. Above mentioned efforts should be more pronounced in immediate future and may extend to other many food legumes like lentil, chickpea which are consumed globally and pigeonpea, blackgram, lathyrus which are having millions of consumers in South-east Asia, particularly Indian subcontinent. The CGIAR centres and NARS of the major food legume growing countries would certainly invest more in legume biofortification programs to harvest immediate gain in securing nutritional food security in vulnerable Asia and African regions in days to come.

Nutrigenomics has a future potential role in detailing out the human nutrition from a different point of view, particularly using the modern genomics tools the entire metabolic interactions of a particular nutrient can be traced out (Fig. 10.1) in vivo as well as in vitro.

Although the potential of this technology as a research tool is promising, the cost to implement these technologies might be a barrier for resource poor countries and regions. Development of low cost research tools, including chemical phenotyping tools, may be necessary to realize global pulse biofortification/bioavailability goals.

Pulses are medium energy and high protein crops that contain a range of micronutrients. Pulse crop development may provide a whole food solution to developing country micronutrient deficiencies as well as a means to reduce the prevalence of diseases of higher-income populations related to high caloric intake. Pulses such as lentil cook quickly (10 min) and thus reduce energy demands for food preparation. In addition, pulses are part of the legume family that fixes atmospheric nitrogen and reduces nitrogen fertilizer demands. Therefore, pulse crop development aimed at nutrient biofortification could provide additional benefits with regards to sustainable development. However, these benefits can only be achieved through research, increasing the acreage of biofortified pulses, and educating consumers regarding the nutritional value of pulse crops. To this end, targeted pulse crop development in major pulse producing and consuming countries is essential. Pulse crops show promise, but delivering upon this promise requires research and resources at greater levels than what is currently being allocated to these crops.

Acknowledgements The North Dakota State University (NSDU) Pulse Quality and Nutrition Laboratory is supported by NDSU Agriculture Experimental Station, the Northern Pulse Growers Association, the USA Dry Pea Lentil Council, and the Bill and Melinda Gates Foundation/HarvestPlus. The authors' previous Canadian pulse research was supported by the Province of Saskatchewan Agriculture Development Fund and the Saskatchewan Pulse Growers Association. The last author thanks Indian Council of Agricultural Research (ICAR) for awarding ICAR International Fellowship.

References

Amarakoon D, Thavarajah D, McPhee K, Thavarajah P (2012) Iron-, zinc-, and magnesium-rich field peas (*Pisum sativum* L.) with naturally low phytic acid: A potential food-based solution to global micronutrient malnutrition. J Food Comp Anal 27:8–13

Ariza-Nieto M, Blair MW, Welch RM, Glahn RP (2007) Screening of iron bioavailability patterns in eight bean (*Phaseolus vulgaris* L.) genotypes using the Caco-2 cell in vitro model. J Agric Food Chem 55:7950–7956

Baum M, Hamwieh A, Furman Bonnie, Sarker A, Erskine William (2008) Biodiversity, genetic enhancement and molecular approaches in lentils. In: Sharma AK, Sharma A (eds) Plant genome, biodiversity and evolution. Enfield NH, Science Publishers, 2008, pp 135–152

Beebe S, Gonzalez AV, Renfigo J (2000) Research on trace minerals in common bean. Food Nutr Bull 21:387–391

Bhatty RS (1988) Composition and quality of lentil (*Lens Culinaris* Medik) – a review. Can I Food Sci Tech J 21:144–160

Broadley MR, White PJ, Hammond JP, Zelko I, Lux A (2007) Zinc in plants. New Phytol 173:677–702

Blair MW, Astudillo C, Beebe S (2005) Analysis of nutritional quality traits in an andean recombinant inbred line population. Bean Improv Coop 48:52–53

Blair MW, Knewtson SJB, Astudillo C, Li C, Fernandez AC, Grusak MA (2010a) Variation and inheritance of iron reductase activity in the roots of common bean (*Phaseolus vulgaris* L.) and association with seed iron accumulation QTL. BMC Plant Biol 10:215

Blair MW, Medina JI, Astudillo C, Rengifo J, Beebe SE, Machado G, Graham R (2010b) QTL for seed iron and zinc concentration and content in a Mesoamerican common bean (*Phaseolus vulgaris* L.) population. Theor Appl Genet 121:1059–1070

Bueckert RA, Thavarajah D, Thavarajah P, Pritchard J (2011) Phytic acid and mineral micronutrients in field-grown chickpea (*Cicer arietinum* L.) cultivars from western Canada. Eur Food Res Technol 233:203–212

Cichy KA, Caldas GV, Snapp SS, Blair MW (2009) QTL analysis of seed iron, zinc, and phosphorus levels in an andean bean population. Crop Sci 49:1742–1750

Combs GF Jr (2001) Selenium in global food systems. Br J Nutr 85:517–547

Dietary Reference Intakes (DRIs) (2012) Food and Nutrition Board, Institute of Medicine, National Academies. https://www.nap.edu. Accessed 15 Oct 2012

Ellis DR, Salt DE (2003) Plants, selenium and human health. Curr Opin Plant Biol 6:273–279

FAOSTAT (2010) http://faostat.fao.org/site/567/default.aspx#ancor. Accessed 10 Sept 2012

Garcia-Casal MN, Leets I, Layrisse M (2000) Ý-carotene and inhibitors of iron absorption modify iron uptake by caco-2 cells. J Nutr 130:5–9

Gawalko E, Garrett RG, Warkentin T, Wang N, Richter A (2009) Trace elements in Canadian field peas: a grain safety assurance perspective. Food Addit Contam A 26:1002–1012

Gelin JE, Forster S, Grafton KF, McClean PE, Rojas-Cifuentes GA (2007) Analysis of seed zinc and other minerals in a recombinant inbred population of navy bean (*Phaseolus vulgaris* L.). Crop Sci 47:1361–1366

Graham RD, Welch RM (2000) Plant food micronutrient composition and human nutrition. Commun Soil Sci Plant Anal 31:1627–1640

Ghandilyan A, Vreugdenhil D, Aarts MGM (2006) Progress in the genetic understanding of plant iron and zinc nutrition. Physiol Plant 126:407–417

HarvestPlus (2012) https://www.harvestplus.org. Accessed 12 Oct 2012

Römheld V, Schaaf G (2004) Iron transport in plants: future research in view of a plant nutritionist and a molecular biologist. Soil Sci Plant Nutr 50:1003–1012

Sarker A, El-Askhar F, Uddin MJ, Million E, Yadav NK, Dahan R, Wolfgang P (2007) Lentil improvement for nutritional security in the developing world. In: Paper presented at the ASA-CSSA-SSSA international annual meetings, New Orleans, LA, 4–8 Nov 2007

Sankaran R, Huguet T, Grusak M (2009) Identification of QTL affecting seed mineral concentrations and content in the model legume *Medicago truncatula*. Theor Appl Genet 119:241–253

Sompong U, Somta P, Raboy V, Srinives P (2012) Mapping of quantitative trait loci for phytic acid and phosphorus contents in seed and seedling of mungbean (*Vigna radiata* (L.) Wilczek). Breed Sci 62:87–92

Talukder ZI, Anderson F, Miklas PN, Blair MW, Osorno J, Dilawari M, Hossain KG (2010) Genetic diversity and selection of genotypes to enhance Zn and Fe content in common bean. Can J Plant Sci 90:49–60

Thavarajah D, Ruszkowski J, Vandenberg A (2008) High potential for selenium biofortification of lentils (*Lens culinaris* L.). J Agric Food Chem 56:10747–10753

Thavarajah D, Thavarajah P (2012a) Evaluation of chickpea (*Cicer arietinum* L.) micronutrient composition: biofortification opportunities to combat global micronutrient malnutrition. Food Res Int 49:99–104

Thavarajah D, Thavarajah P (2012b) Lentils (*Lens culinaris* L): a superior whole food solution to micronutrient malnutrition. In: Betancourt A, Gaitan H (eds) Micronutrients sources, properties, and health effects. Nova Science, New York, pp 179–191

Thavarajah D, Thavarajah P, Sarker A, Vandenberg A (2009) Lentils (*Lens culinaris* Medikus Subspecies *culinaris*): a whole food for increased iron and zinc intake. J Agric Food Chem 57:5413–5419

Thavarajah D, Thavarajah P, Wejesuriya A, Rutzke M, Glahn RP, Combs GF, Vandenberg A (2011) The potential of lentil (*Lens culinaris* L.) as a whole food for increased selenium, iron, and zinc intake: preliminary results from a 3 year study. Euphytica 180:123–128

Viadel B, Barberí R, Farr R (2006) Uptake and retention of calcium, iron, and zinc from raw legumes and the effect of cooking on lentils in Caco-2 cells. Nutri Res 26:591–596
Waters BM, Grusak MA (2008) Quantitative trait locus mapping for seed mineral concentrations in two *Arabidopsis thaliana* recombinant inbred populations. New Phytol 179:1033–1047
Walker DR, Scaboo AM, Pantalone VR, Wilcox JR, Boerma HR (2006) Genetic mapping of loci associated with seed phytic acid content in CX1834-1-2 Soybean. Crop Sci 46:390–397
Wang WC, Nanto V, Makela AL, Makela P (1995) Effect of nationwide selenium supplementation in Finland on selenium status in children with juvenile rheumatoid arthritis. A ten-year follow-up study. Analyst 120:955–958
Welch RM (2002) Breeding strategies for biofortified staple plant foods to reduce micronutrient malnutrition globally. J Nutr 132:495S–499S
Welch RM, Graham RD (1999) A new paradigm for world agriculture: meeting human needs: productive, sustainable, nutritious. Field Crops Res 60:1–10
Welch RM, Graham RD (2004) Breeding for micronutrients in staple food crops from a human nutrition perspective. J Exp Bot 55:353–364
White PJ, Broadley MR (2009) Biofortification of crops with seven mineral elements often lacking in human diets – iron, zinc, copper, calcium, magnesium, selenium and iodine. New Phytol 182:49–84
Wu L, Guo X, Banuelos GS (1997) Accumulation of seleno-amino acids in legume and grass plant species grown in selenium-laden soils. Environ Toxicol Chem 16:491–497
Yeung CK, Glahn RP, Welch RM, Miller DD (2005) Prebiotics and iron bioavailability-is there a connection? J Food Sci 70:R88–R92

Chapter 11
Towards Enriching Genomic Resources in Legumes

Aditya Pratap, Rakhi Tomar, Neha Rajan, Jitendra Kumar, Pooja Bhatnagar Mathur, Nupur Malviya, and Tuba K. Anjum

Abstract Food legumes, mainly comprising dry beans, dry peas, soybean, chickpea, pigeonpea, groundnut, greengram, blackgram, cowpea, lentil and lathyrus, have considerable area under cultivation globally and these are important constituents of cereal-based vegetarian diets. Keeping in view their tremendous importance for diversification and intensification of contemporary agriculture, systematic efforts towards their genetic improvement have been undertaken with classical breeding tools, lately complemented by the use of genomic tools. These genomic tools provide comprehensive information on genes involved in biochemical pathways leading upto nutritional compounds and can be used to understand the genetics of traits of interest and consequently, helping in marker assisted breeding. During the last two decades powerful genetic and genomic tools such as establishment of genetic and physical maps, expressed sequence tags, bioinformatic tools, genome-wide sequence data, genomic and metabolomic platforms, etc. have been developed for many legume species. These efforts have led to development of large scale molecular markers, identification of various marker trait associations, construction of genetic and linkage maps, expressed sequence tags database, partial or whole genome sequences, physical and molecular maps, DNA chips and bacterial artificial chromosome (BAC) libraries. After the genome sequencing of three model species, *Medicago*, *Lotus* and *Glycine*, draft genome sequences have recently been made

A. Pratap, Ph.D. (✉) • R. Tomar, M.Sc. • J. Kumar, Ph.D. • N. Malviya, Ph.D.
T.K. Anjum, M.Sc.
Crop Improvement Division, Indian Institute of Pulses Research,
G.T. Road, Kanpur, Uttar Pradesh 208024, India
e-mail: adityapratapgarg@gmail.com

N. Rajan, M.Sc.
Krishi Vigyan Kendra, Ramakrishna Mission Ashram, Ranchi, Jharkhand, India

P.B. Mathur, Ph.D.
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT),
Hyderabad, Andhra Pradesh, India

S. Gupta et al. (eds.), *Legumes in the Omic Era*, DOI 10.1007/978-1-4614-8370-0_11,

available in agronomically important food legumes, pigeonpea and chickpea while similar efforts are underway in groundnut and greengram. The new generation sequencing (NGS) and genotyping platforms such as 454/FLX sequencing and Illumina GoldenGate/Solexa have revolutionized plant genomic research as these generate millions of ESTs per run. With the increased amount of genomic resources, there are now tremendous opportunities to integrate these with the genetic resources for their widespread use in routine legume improvement programmes by integrating them with conventional breeding tools. As a result, the genomics assisted breeding (GAB) can now be successfully used in legume improvement and development of improved genotypes having improved agronomic and quality traits and resistance to biotic and abiotic stresses. This chapter discusses the developments made in development of legume genomics and their role in overall improvement of food legumes.

Keywords Genomic resources • Molecular markers • Genomic library • Whole genome sequencing • Comparative genomics • Genomics assisted breeding

Introduction

Legumes are important source of food, feed and fodder in many agricultural systems and are grown on a large scale in semi-arid tropics of the world. Grain legumes alone contribute 33 % of human protein nutrition (Vance et al. 2000) and have a unique ability to fix the atmospheric nitrogen in symbiotic association with *Rhizobium* bacteria, which not only enables them to meet their own nitrogen requirement but also benefit the succeeding crops. Improvement in agronomic and phenological traits of the legumes is crucial in order to improve their use as human food and sustainability of production system. Therefore, yielding ability, seed and quality characteristics, resistance to biotic and abiotic stresses, storability, etc. are receiving greater attention for the genetic improvement of legumes. There is also an increasing interest in improving nutritional characteristics of legumes with enhanced content of β-carotene, leutin, isoflavones and other nutraceuticals.

The way to development of better food and forage legumes requires a detailed knowledge of the different genes involved in biochemical pathways leading upto nutritional compounds, including the expression patterns and level of these genes and their interactions (Gepts et al. 2005). Genomic resources are important to understand the genetics of traits of interest and consequently, marker assisted backcross breeding (MABC), marker assisted recurrent selection (MARS) and advanced backcross (AB) breeding may be used effectively in legume improvement. A great success in this will be possible by combining genomic tools with rational selection of germplasm and precise phenotyping for traits of interest, termed as "genomics-assisted breeding" (Varshney et al. 2005). During the last two decades powerful genetic and genomic tools such as establishment of genetic and physical maps, expressed sequence tags (ESTs), bioinformatic tools, genome-wide sequence data,

genomic and metabolomic platforms have been developed for many legume species. This chapter gives a comprehensive view of development and utilization of genomic resources in major food legume crops.

Genomic Resources in Legumes

Over the past many years, there has been an increased focus on application of powerful genomic approaches to major legume species with an aim of generating genomic resources that will not only be of use in these species but also facilitate crop improvement in other species also. Apart from two model legumes, *Medicago truncatula* and *Lotus japonicus*, efforts have been made in developing genomic resources in common bean (*Phaseolus vulgaris*), cowpea (*Vigna unguiculata*), soybean (*Glycine max*), pigeonpea (*Cajanus cajan*), *alfa alfa* (*Medicago sativa*), chickpea (*Cicer arietinum*), faba bean (*Vicia faba*), lentil (*Lens culinaris*), pea (*Pisum sativum*) and peanut (*Arachis hypogaea* L.). However, these legumes differ greatly in their genome size, base chromosome number, ploidy level, and compatibility status (Table 11.1). The efforts have led to development of large scale molecular markers, identification of various marker trait associations, construction of genetic and linkage maps, expressed sequence tags (EST) database, partial or whole genome sequences, physical and molecular maps, DNA chips and bacterial artificial chromosome (BAC) libraries in all these crops (Table 11.2 and 11.3). Among the agronomically important food legumes, draft genome sequence has recently been made available in pigeonpea (Singh et al. 2012; Varshney et al. 2012) and chickpea (Varshney et al. 2013a) and similar efforts are underway in groundnut.

Table 11.1 Variation in basic chromosome number and genome size among legume species

Species	Basic chromosome number (X)	Genome size (Mb)
Peanut (*Arachis* spp.)	10–20	1,260–2,890
Lupin	5–13	468–1,177
Common bean (*Phaseolus vulgaris*)	11	637
Cowpea (*Vigna unguiculata*)	11	620
Pigeonpea (*Cajanus cajan*)	11	858
Soybean (*Glycine max*)	20	1,115
Lotus japonicus	6	472
Pea (*P. sativum*)	7	4,400
Lentil (*L. culinaris*)	7	4,063
Chickpea (*C. arietinum*)	8	740
Alfalfa (*M. sativa*)	8	800–900
Vicia faba	7	–
Medicago truncatula	8	500–550

Table 11.2 Current availability status of genomic resources in pulses

Genomic resources	Chickpea	Common bean	Pigeonpea	Cowpea	Lentil	Mung/Urdbean	Fieldpea
Mapping populations	+	+	+	+	+	+	+
BAC based resources							
BAC libraries	+	+	+	+	–	+	+
BAC-end sequences	+	+	+	+	–	–	–
Physical map	+	+	*	+	–	–	–
Second and third generation DNA markers							
Genomic SSRs	+	+	+	+	+	+	+
Genic or EST-SSRs	+	+	+	+	+	+	+
SNP	+	+	+	+	–	–	+
DArT	+	+	+	–	–	–	
SFP	–		+	+	–	–	–
Transcriptomic resources							
ESTs	+	+	+	+	+	+	+
Transcriptome assemblies	+	+	+	+	+	+	+
Published genetic maps							
Population specific (inter-specific/inter-subspecific and cultivated)	+	+	+	+	+	+	+
Consensus maps	+	+	+	+	–	–	+
Whole genome sequence	+	*	+	*	*	*	*

+ available; – not available; *in progress

Table 11.3 Important genomic resources in major food legumes developed in last 5 years

Genomic resources	Crop	References
BAC libraries and BAC end sequences	Chickpea	Thudi et al. (2011)
	Common bean	Córdoba et al. (2010)
	Pigeonpea	Bohra et al. (2011)
	Cowpea	Yu (2012); http://www.comparative-legumes.org/pages/resources
Large scale SSR/SNP markers	Chickpea	Thudi et al. (2011); Hiremath et al. (2012); Gaur et al. (2012)
	Common bean	Hyten et al. (2010)
	Pigeonpea	Raju et al. (2010); Bohra et al. (2011); Dubey et al. (2011); Kassa et al. (2012)
	Cowpea	Muchero et al. (2009); Lucas et al. (2011)
High throughput genotyping platforms		
DArT arrays	Chickpea	Varshney et al. (2010)
	Common bean	Briñez et al. (2011)
	Pigeonpea	Yang et al. (2011)
GoldenGate/KASPar assays	Chickpea	Hiremath et al. (2012); Gaur et al. (2012)
	Common bean	Cortés et al. (2011); Hyten et al. (2010)
	Pigeonpea	Kassa et al. (2012)
	Cowpea	Muchero et al. (2009); Lucas et al. (2011)
First genetic maps	Mungbean	Isemura et al. (2012)
	Pigeonpea	Yang et al. (2011); Bohra et al. (2011, 2012)
High density genetic maps	Chickpea	Thudi et al. (2011), Hiremath et al. (2012), Gaur et al. (2012)
	Common bean	Galeano et al. (2011)
	Cowpea	Muchero et al. (2009); Lucas et al. (2011)
Physical maps	Chickpea	Zhang et al. (2010)
	Common bean	http://cmap.comparative-legumes.org
	Cowpea	http://phymap.ucdavis.edu/cowpea/
Draft genome sequences	Pigeonpea	Singh et al. (2012); Varshney et al. (2012)

Genome Sequences

The three model species, *Medicago*, *Lotus* and *Glycine* are the first legume crops to have their genomes sequenced. Among these, *M. trancatula* and *L. japonicus* were chosen for genome sequencing largely because of their small diploid genomes (ca. 500 and 471 Mb in size), shorter life cycle and availability of supportive resources (Young et al. 2005). The information generated by genome sequencing of these two species has provided greater insight into their gene structure as well as their physical and genetic maps. Though, the sequencing of both these species was initiated at almost the same time, the approaches used for sequencing differed slightly in these. While for sequencing of *Lotus* genome, a modified BAC-by-BAC approach followed by draft sequencing of the selected regions of the genome was followed, in *Medicago* genome sequencing project,

a traditional BAC-by-BAC approach was followed, though it was focused on the euchromatic part of the genome. In *Medicago*, 0.6–0.7 of the estimated euchromatic genomic region has been sequenced, capturing about 0.60 of the genes (Kumar et al. 2011). The sequencing is expected to be completed soon; having an assembly of c. 300 Mb and capturing about 0.90 % of the genes. In case of *Lotus* genome also, considerable progress has been made with sequencing of about 0.67, covering 0.91 % of the gene space (Sato et al. 2008). In both cases, however, the traditional type of sequencing method was used.

For sequencing of soybean, a Phaseoloid legume, another genome sequencing method- whole genome shotgun (WGS)- was used. Soybean is an excellent representative of polyploid species and it was chosen as a model legume for sequencing (Gepts et al. 2005) due to its moderate genome size (ca. 1,115 Mb), available infrastructure (Jackson et al. 2006) and also due to its economic importance (Nunberg et al. 2006). Soybean WGS comprises 950 Mb of assembled and anchored sequences representing about 0.85 of the predicted genome size. It has been predicted that the soybean genome has 46,430 protein coded genes and about 0.75 of these genes are there in multiple copies (Schmutz et al. 2010). Though this approach is powerful and fast, but it is largely suitable to smaller and less complex genomes. Another Phaseoloid legume, common bean, a diploid species, has medium sized genome (588–637 Mb) (Bennett and Leitch 2012). Besides the small size, it was chosen for genome sequencing due to availability of good amount of genomic resources such as availability of 9X physical map, BAC libraries, 25 linkage, 83,530 ESTs and knowledge of the genic (0.29) and repetitive (0.49) portions of the genome (see Kumar et al. 2011). Its extensive macrosyntentic relationships with soybean has also favoured its candidature for best model species for soybean and other legume species in order to develop new SSR and SNP markers and also for identification of candidate genes.

Most recently, the draft genome sequence has been made available in chickpea, the second most important grain legume after soybean. In *kabuli* chickpea variety, CDC Frontier, ~738 Mb long draft WGS sequence has been reported which contains 28,299 genes (Varshney et al. 2013a). Re-sequencing of 90 more chickpea genotypes was also done which provided an access to millions of genetic markers and low diversity genome regions that may be useful in the development of superior varieties with enhanced drought tolerance and disease resistance. The genome map will also help tremendously in harnessing genetic diversity by broadening the genetic base of cultivated chickpea genepool. In pigeonpea, draft genome sequence has been made available by two independent groups almost at the same time (Singh et al. 2012; Varshney et al. 2012). For generating the genome sequence in this crop, the ICRISAT led team used Illumina next-generation sequencing platform to generate 237.2 Gb of sequence, which along with Sanger-based bacterial artificial chromosome end sequences and a genetic map, was assembled into scaffolds representing 72.7 % (605.78 Mb) of the 833.07 Mb pigeonpea genome. Genome analysis predicted 48,680 genes for pigeonpea and also showed the potential role that certain gene families have played throughout the domestication of pigeonpea and the evolution of its ancestors. In another independent approach by Singh et al. (2012),

the whole genome of pigeonpea was assembled using long sequence reads of 454 GS-FLX sequencing with mean read lengths of >550 bp and >10X genome coverage, resulting in 510,809,477 bp of high quality sequence. Total 47,004 protein coding genes and 12,511 transposable elements related genes have been predicted in this study. Further, 1,213 disease resistance/defense response genes and 152 abiotic stress tolerance genes were also identified. This genome sequence was also used to identify large number of hypervariable pigenpea simple sequence repeat (HASSR) markers, 437 of which have been experimentally validated for PCR amplification and high rate of polymorphism among pigeonpea varieties. These markers will be immensely useful for fingerprinting and diversity analysis of pigeonpea germplasm and molecular breeding applications. Efforts are already underway to make the draft genome sequence available in peanut very soon. However, in most of the other food legumes, with the exception of pea (*P. sativum*), alfalfa (*M. sativa*), peanut (*Arachis hypogaea*) and cowpea where some progress has been made recently, lesser genomic information is available. In cowpea, genome filtering method has been used for sequencing and analyzing the gene-rich regions (hypomethylated portion of the cowpea genome). This has led to development of >250,000 gene-space sequence reads (GSRs) with an average length of 610 bp yielding ~160 Mb of sequence information (Timko et al. 2008). Among the GSR dataset, 29 % of the sequences annotated using the *Arabidopsis* gene ontology (GO) was involved to encode the majority of cellular enzymes and components of amino acid, carbohydrate and lipid metabolism. Besides, a total of 5,888 GSRs had homology to genes encoding transcription factors (TFs) and about 5 % of the total annotated sequences in the dataset have represented transcription associated factors (TAFs). This information can be utilized in mapping and tagging the genes for agronomically important traits in legumes.

BAC/BIBAC Resources

The bacterial artificial chromosome (BAC) and binary bacterial artificial chromosome (BIBAC) libraries are good genomic resources that allow genome sequencing, development of new molecular markers and physical map, and map based cloning of genes (Tao et al. 2001). In several legume species, these libraries have been developed with varying clone sizes from 100 to 150 kb. In chickpea, a BIBAC library of 23,780 clones, with an average insert size of 100 kb and a coverage of 3.8 genome equivalents, was prepared for facilitating the development of not only genomic SSRs but also gene specific SSRs (Rajesh et al. 2004).

In soybean, a genome-wide physical map has been constructed from more than 78,000 BAC clones, representing 9.6X genome. It consisted of approximately 2,905 contigs which were estimated to span 1,408 Mb in physical length (Wu et al. 2004). More than half of the length of the physical map was anchored to the genetic map using 388 DNA markers. Earlier, using molecular markers, physical map from BAC clones could also be related to the genetic map by locating existing genetic markers on the contigs (Lewers et al. 2002). These contigs work as a starting point for

positional cloning of specific genes which has accelerated the discovery of genes underlying phenotypes of agronomic interest (Liu et al. 2001; Xu et al. 2001). BAC libraries have also been used to generate SSR markers leading to identification of two genomic regions involved in resistance to the soybean cyst nematode (Cregan et al. 1996, 1999). Moreover, these also helped in fine mapping of genes leading to identification of tightly linked markers for marker-assisted selection. In cowpea, 60,000 BAC clones were assembled into a 10X physical map and efforts are already underway to anchor the cowpea physical map to the emerging SNP-based genetic linkage map. In common bean, sequencing of 89,000 BAC ends has yielded a 9X draft physical map which represents 62 Mb of genome sequence or 9.5 % of the common bean genome (Schlueter et al. 2008). In this map, 540 markers derived from RFLPs, genes, ESTs and other sequences have been anchored, of which 84 are genetically mapped and provide linkage between the physical and genetic maps (see at http://phaseolus.genomics.purdue.edu/).

In pea (*Pisum sativum* L.), two BAC libraries, which are useful resources for the isolation of genes underlying disease resistance and other economically important traits have been constructed. These libraries separately contained 55,680 and 65,280 clones, of which ~1 % clones were from chloroplast origin (Coyne et al. 2007). In peanut also, the BAC libraries from the AA genome (*Arachis duranensis*) with 84,096 clones and from the BB genome (*A. ipaënsis*) with 75,648 clones having average insert size of 110 and 100 kb, have been constructed. An estimate based on the library average insert size and *A. duranensis* haploid genome equivalent to 1,260 Mb showed that the coverage of the AA genome BAC library is equivalent to 7.4X genome. However, for *A. ipaënsis*, the DNA-content determination is controversial and hence the BB genome BAC library for *A. ipaënsis* could represent from 2.7 to 5.3 the haploid genome equivalents of the species considering the earlier discrepancies in estimation of haploid genome size (Varshney et al. 2009b). The BAC-based resources developed in different species will have greater utility for subsequent genome analyses, because they provide the basis for a physical interpretation of other genetic and genomic resources within each species, and they will facilitate more detailed analysis of high value regions of the genomes of legumes.

Molecular Markers

In most of the legume species, several DNA-based marker systems such as single nucleotide polymorphism (SNP), random amplified polymorphic DNA (RAPD), simple sequence repeats (SSRs) or microsatellites, amplified fragment length polymorphisms (AFLPs) and hybridization based marker systems such as restriction fragment length polymorphisms (RFLPs) and diversity arrays technology markers (DArT) are now available. However, PCR (Polymerase Chain Reaction) based SSR and SNP markers are preferred by breeders because of their high reproducibility, high level of polymorphism and user friendliness. SSR markers have the advantage of being multi-allelic and co-dominant (Gupta and Varshney 2000). Further, these can easily be employed in genotyping of large segregating populations in a

cost-effective manner and with minimum infrastructure facilities. While in many crops, these have been extensively utilized, in pulses their use is still limited to only a few crops like chickpea and pigeonpea (Varshney et al. 2009a; Saxena and Nadarajan 2010). Among the different marker systems used in pulses and other crop plants, SNP markers are high throughput and cost effective (Varshney et al. 2012). Similarly, diversity array technology (DArT) marker system is used for diversity studies, saturating linkage maps and identifying alien introgressions. The following section describes the most popular molecular marker systems in legumes.

SSRs

Since the SSR markers are the markers of choice in legume improvement, their availability has great significance in legume species for practical purposes. Over the years, a large number of SSR makers have been developed for many legume species by using following approaches individually or in combination (Varshney et al. 2012): (a) construction and sequencing of SSR enriched genomic DNA libraries, (b) sequencing and mining the BAC (bacterial artificial chromosome)-end sequences (BES) for SSRs, and (c) mining the transcript sequences generated by either Sanger sequencing or next generation sequencing (NGS) approaches such as 454/FLX sequencing (for details see Kumar et al. 2011). Most recently, 487 novel markers including 125 EST-SSRs, 102 SNPs, 151 intron targeted primers, 109 EST polymorphisms have been developed in chickpea (Choudhary et al. 2012). Similarly, about 2,000 new SSRs have also been developed earlier using genomic DNA libraries (Nayak et al. 2010; Gaur et al. 2011), ESTs (Varshney et al. 2009b), BAC end sequences (Thudi et al. 2011) and 454/FLX transcript reads (Garg et al. 2011a, b). These markers are also in use in other legume species including cowpea (768 BAC end sequence-BES-SSRs), lentil (100 genomic SSRs) and common bean (ca. 500 SSRs) (see Kumar et al. 2011). In peanut, ca. 6,000 markers are now available for use (Pandey et al. 2012). Most recently, 3,072 BES-were developed in pigeonpea (Bohra et al. 2011). Besides 3,583 SSRs from ESTs (Raju et al. 2010) and 454/FLX sequences (Dubey et al. 2011 and Dutta et al. 2011) are also available for molecular marker assisted breeding programmes.

DArT

DArT marker system has tremendous use for diversity studies and identification of alien introgressions, especially from wild species into the cultivated ones. Recently, by using 1225 DArT markers in the cross between *C. platycarpus* and *C. cajan*, 2–5 % *C. platycarpus* genome-carrying genes for disease and insect resistance were observed (Mallikarjuna et al. 2011). Yang et al. (2011) developed first generation array comprising 6,144 clones in pigeonpea. Similarly, ICRISAT has developed

Table 11.4 EST database of food legumes (as on 28 Nov 2012 at NCBI)

Common name	Botanical name	EST submitted in NCBI
Soybean	*Glycine max*	1468424
Burclover	*Medicago truncatula*	286175
Cowpea	*Vigna unguiculata*	189593
Chickpea	*Cicer arietinum*	46064
Pigeonpea	*Cajanus cajan*	25577
Mungbean	*Vigna radiata*	1604
Blackgram	*Vigna mungo*	311
Field pea	*Pisum sativum*	21837
Common bean	*Phaseolus vulgaris*	149769

DArT assays comprising 15,360 clones in chickpea, pigeonpea and groundnut and diversity study using these showed a narrow genetic diversity in the elite gene pool in comparison to the landraces and wild species (Varshney et al. 2012). Recently, DArT arrays have also become available in common bean (Briñez et al. 2011).

EST Databases

Extensive efforts have been made in sequencing expressed genomic regions obtained from tissues in different conditions and developmental stages, leading to deposition of large number of EST sequences in the public database (Kumar et al. 2011; Table 11.4). The EST databases provide an effective tool for gene discovery and generate raw material for the production of cDNA arrays for transcriptome analysis (Coram and Pang 2005a). As a result, these easily accessible EST sequences have emerged as cost-effective valuable source for in silico generation of markers and broaden the field of comparative mapping in species where limited or no sequence information is available. EST database provides the first insight into the genes that may be associated with root development and abiotic stress tolerance, particularly in crops like chickpea (Jayashree et al. 2005). EST libraries have been generated and analysed in chickpea for isolation of candidate genes controlling defence mechanism in *Ascochyta* blight (Coram and Pang 2005b). The identified ESTs have putative relationships with proteins involved in drought tolerance and hence provided a useful resource for identification of candidate gene or mining the alleles responsible for drought avoidance and tolerance in cool season legumes (Buhariwalla et al. 2005).

Microarrays or DNA Chips and Transcriptome Analysis

Microarrays or DNA Chips are important tools of functional genomics for identifying the network of genes underlying the expression of agronomically important traits (Meyers et al. 2004). These can be developed from hundreds of thousands ESTs or cDNA libraries available in model and other legume species (Table 11.5).

Table 11.5 Microarray data tools developed in legume species

Species	No. of genes/cDNA clones on microassays	Plant tissue from which cDNA library developed	Remarks	References
M. truncatula	2,268 cDNA	Arbuscular mycorrhiza root	–	Liu et al. (2003)
	6,000 unigenes	Root symbiotic interaction	Mt6k-RIT	Yahyaoui et al. (2004)
	2,26,923 high quality ESTs	The TIGR Gene Index databases	–	Lee et al. (2005)
	Mt6k-RIT clones as well as 6,144 unigenes	Different stage of root nodule development	–	Lohar et al. (2006)
	Mt6k-RIT extended with 2 k-set of cDNA clones	Flowering and pod development stage	Mt8k-RIT	Firnhaber et al. (2005)
	16,086 oligo probes from tentative consensus (TC) sequences	Different developmental tissues	Mt16kOLI1	http://www.eugrainlegumes.org
	Mt16LI1 oligo probes along with 384 probes targeting TFs and other regulators	Different developmental tissues	Mt16kOLI1Plus (This microarray chips represent 35–40 % genes of *M truncatula*	http://www.genetik.uni-bielefeld.de/MolMyk/
Pea	70mer oligonucleotide microarray derived from ~5,200 EST	Cotyledon and seed coat ESTs	Ps6kOLI1	http://www.grainlegumes.com/aep/r_d_projects/grain_legumes_glip/progress_in_glip/integrated_activities/gene_expression_profili1ng
Mungbean	EST 6,272 genes (a unigene set of 2,013 comprising of 973 contigs and 1,040 singletons and 262 genotype specific SNPs)	Etiolated seedlings	–	Kuhar et al. (2012)
Chickpea	2,100 EST 65,000–68,000 singletons	Root and collar tissue from 25 days old seedlings	–	Ashraf et al. (2009)
Phaseolus	2,900 cDNA clones	–		Gepts et al. (2008)

In *M.truncatula*, microarrays developed using EST and other oligo sequences were used to study the expression of genes involved in nodule formation during symbiotic association, and in development of flower, pod and seed. A commercial affymetrix chip with a 51 k GeneChip including cDNA-microarrays and 70-mer oligonucleotide microarrays of different tissues developed in this species are useful genomic resource for comparative analysis of gene expression in related grain and forage legumes. The use of these DNA microarrays/chips has led to identification of thousands of genes that are induced or repressed during the development of nodules and symbiotic nitrogen fixation (Kuester et al. 2004; Baier et al. 2007; Benedito et al. 2008; Jones et al. 2008). Several other studies have also investigated the transcriptional basis of seed development, differentiation, desiccation, plant responses to aluminium toxicity, and changes in nitrogen nutrition using microarray chips (Buitink et al. 2006; Verdier et al. 2008; Narasimhamoorthy et al. 2007; Chandran et al. 2008). Long-oligo arrays of *M. truncatula* have also been used effectively to identify the transcripts upregulated in alfalfa trichomes secreting the molecules during insect defense (Aziz et al. 2005).

In Pea (*Pisum sativum*), a microarray (Ps6kOLI1) consisting of 70-mer oligo probes targeting ~5,200 EST clusters assembled predominantly from cotyledon under GLIP has been developed primarily for identifying genes relevant to seed formation. Similarly in soybean, high-density expression arrays containing 18,000 cDNAs arrayed on a filter have been developed (Shoemaker et al. 2003) and three microarrays comprising low redundancy unigene sets of 27,513 clones (each microarray with 9,728 unigenes) have been constructed from a variety of cDNA libraries made from a wide range of organs at different developmental stages, disease-challenged tissues, and various stress conditions. These microarrays have been used to examine tissue specific gene expression and global expression in mutant isolines which led to identification of set of candidate genes potentially encoded or modulated by the mutant phenotype (Vodkin et al. 2004). The microarray tools developed in soybean have been used successfully to identify genetic markers closely linked to soybean aphid resistance gene *Rag1* (Kaczorowski et al. 2008), and genes involved in the soybean iron deficiency chlorosis response under iron deficient conditions (O'Rourke et al. 2007).

In chickpea, 768-feature microarray was developed that comprised 559 chickpea cDNAs, 156 grass pea cDNAs, 41 lentil resistance gene analogs (RGAs) and 12 controls. Using this microarray, the transcriptional change in genes responsible for different abiotic stresses was observed leading to identification of 2, 15 and 30 genes differentially expressing between tolerant and susceptible genotypes for drought, cold and high-salinity, respectively. These genes code for various functional and regulatory proteins. Significant differences in stress responses were observed within and between tolerant and susceptible genotypes highlighting multiple gene control and complexity of abiotic stress response mechanism in chickpea (Mantri et al. 2007, 2010). In case of lentil also, a cDNA microarray approach has deciphered the *Ascochyta* blight resistance (Mustafa et al. 2009).

New Generation Tools for Legume Genomics

High Throughput Sequencing/Genotyping Platform

New generation sequencing (NGS) and genotyping platforms such as 454/FLX sequencing and Illumina GoldenGate/Solexa have revolutionized plant genomic research by generating millions of ESTs per run. The advantage with these sequencing methods is that these are not limited by prior knowledge of transcribed sequences or predicted genes. Approximately 75 million ESTs have been generated in *M. truncatula* using an Illumina/Solexa resulting in quantitative expression data complement and extend Affymetrix Gene Chip data (Benedito et al. 2008; Young and Udvardi 2009). Next-generation sequencing may also become an attractive option for transcriptomics of non model species where DNA arrays are unavailable, especially if sequence lengths can be increased to facilitate alignment and contig assembly. Using 454/FLX sequencing at ICRISAT in collaboration with JCVI and NCGR, 435,184 and 496,705 sequence reads providing 44,852 and 48,519 contigs were obtained from chickpea and pigeonpea, respectively. These sequence data provide access to a significant fraction of the total transcriptomes of these crops, and are expected to aid in the analysis of drought tolerance, including candidate gene discovery and the development of molecular markers for breeding applications (Varshney et al. 2005). In another study, 2,496 ESTs were generated and utilized in chickpea for the development of 487 novel EST-derived functional markers including 121 EST-SSRs, 151 intron targeted primers, 109 EST polymorphisms (ESTP) and 102 SNPs (Choudhary et al. 2012). While EST-SSRs, ITPs and ESTPs were developed by in silico analysis of the developed EST sequences, SNPs were identified by allele resequencing and their genotyping was done using Illumina GoldenGate Assay. In groundnut, Sanger sequencing, which is slightly more extensive, has been conducted which resulted in 54,000 ESTs for cultivated groundnut (*A. hypogaea*) and 6,000 in the diploid *A. stenosperma.*

The NGS platforms are also important tools for discovery of SNPs, especially in legumes having a narrow genetic base. Development of large-scale SNP markers may help accelerate linkage mapping and whole genome association (WGA) studies. In this connection, efforts have been made by several institutions for developing the SNP markers in cowpea, pigeonpea, chickpea and groundnut (reviewed by Varshney et al. 2009a, b). Recently, 26,082 SNPs have been identified in chickpea based on alignment of approximately 37 million Illumina/Solexa tags generated from ICC4958 and ICC1882 genotypes (Hiremath et al. 2011). In pigeonpea, 12,141 SNPs have been identified in ten parental genotypes based upon the alignment of 160 million reads against a transcriptome assembly (CcTAversin 1.0) (Dubey et al. 2011). Further, comparison of transcript reads from 12 different pigeonpea genotypes has led to identification of 28,104 novel SNPs (Varshney et al. 2012). Kudapa et al. (2012) developed a comprehensive transcriptome assembly for pigeonpea by

analysing 128.9 million short Illumina GA IIx reads, 2.19 million single FLX/454 reads and 18,353 Sanger expressed sequenced tags from more than 16 genotypes. Based upon the knowledge of intron junctions, 10,009 primer pairs were designed from 5,033 TACs for amplifying intron spanning regions (ISRs). These ISR markers will be immensely beneficial to accelerate breeding and genetic research in pigeonpea. Similarly, KASPar assays from another next generation SNP genotyping technology, have also been developed for 2,005 SNPs in chickpea (Hiremath et al. 2012) and 1,616 in pigeonpea.

Serial Analysis of Gene Expression

Serial analysis of gene expression (SAGE) is an approach that allows rapid and detailed analysis of thousands of transcripts. In case of chickpea, 80,238 26-bp tags representing 17,493 unique transcripts (UniTags) from drought-stressed and non-stressed control roots have been generated using SuperSAGE technology for the analysis of gene expression in chickpea roots in response to drought (Molina et al. 2008). Sanger sequencing has been used to a limited extent to access the chickpea and pigeonpea transcriptomes (27,000 and 13,000 ESTs, respectively).

RNAi and TILLING

Forward genetics which aims at identifying the responsible genes for a trait, can be performed through map based cloning and T-DNA and transpose insertional or insertion mutagenesis. This has been used widely for identification and cloning of genes for a known phenotype (Kumar et al. 2011). For example, in *L. japonicus*, two new *Sym* genes (*LjSym1* and *LjSym2*) have been isolated through map-based cloning approach. *LjSym2* is required for symbiosis involving both arbuscular mycorrhizal (AM) fungi and rhizobia in root nodules (RNs) while, the *LjSym2* gene encodes a receptor-like kinase (Endre et al. 2002). Another important approach is reverse genetics approach for which mutant population can be a valuable resource. Such mutant populations can be generated through T-DNA and retrotransposon insertions where gene sequences or a protein with unknown function are associated with responsible phenotype. Following this approach, a population in *M. truncatula* mutagenized by a tobacco retrotransposon, *Tnt1* has become an important resource for reverse genetics (D'Erfurth et al. 2003). Screening this population by sequencing of tagged sites led to the isolation of *M. truncatula"Pim"*gene (Benlloch et al. 2006).

More recently, RNAi technology or virus induced gene silencing, have become important resources for knowing the function of genes (Allen et al. 2004). In legumes, virus-induced gene silencing has been used in pea (Constantin et al. 2004). In soybean, RNAi induced gene silencing has been successful using transformation methods, either through biolistics or *Agrobacterium tumefaciens* (Reddy et al. 2003;

Subramanian et al. 2005; Nunes et al. 2006), hairy root transformation (Jackson et al. 2006), transposon mutagenesis (Jackson et al. 2006) and virus-induced gene silencing. Targeting induced local lesions in genomes (TILLING) or deletion-TILLING (de-TILLING) is a reverse genetics approach which uses knowledge of gene sequence having unknown function to know their function or phenotype. A large number of TILLING resources have been developed in several legume species (Table 11.6). Using this approach, approximately 2,000 individual germlines have been generated in *Medicago truncatula* (Vanden Bosch and Stacey 2003). Similarly, in *Lotus japonicusa* population of >40,000 mutants was developed through induced mutation by using 1 % v/v EMS comprising mutants defective for morphological, metabolic and nodule formation characters (Perry et al. 2003) and also the mutants having variant alleles of SYMRK and sucrose synthesis genes using TILLING procedure (Stracke et al. 2002; Horst et al. 2007). The TILLING resources developed in different legumes have provided notable functional genomic resources to the legume researchers towards knowing the function of genes.

Use of Genomic Resources in Legume Improvement

With the development of large scale genomic resources in major food legumes, there are now tremendous opportunities to integrate them with genetic resources for their widespread use in routine breeding practices and their integration with conventional breeding tools. As a result, the genomics assisted breeding (GAB) can now be successfully used in legume improvement for development of improved genotypes having resistance to biotic and abiotic stresses and improved agronomic traits. The available genomic resources have successfully been used in legumes for hybridity confirmation, diversity analysis studies, marker assisted breeding, genome wide selection and advanced back cross QTL analysis.

Hybridity Confirmation

In most of the legumes species, making crosses is difficult as compared to cereals owing to small size of the flower and a weak peduncle supporting the bud. Legumes being self pollinated crops, have increased chances of selfing. Furthermore, differentiating between the selfed and F_1 plants is also difficult due to low phenological diversity between the selfed and crossed plants. Marker assisted identification of true F_1 hybrids is a robust and full-proof approach for identification of true hybrids and therefore increasing the efficiency of selection of desired recombinants. This approach is now being routinely used in identification of true F_1 plants in chickpea in the crosses between Pusa 256×Vijay and Pusa 256×WR315 at IIPR, Kanpur; C104×WR315 and C 214×ILC 3279 at ICRISAT; JG 74×WR 315 at JNKVV, Jabalpur; Phule G12×WR 315 at MPKV, Rahuri and Annigeri-1×WR 315 at ARS

Table 11.6 TILLING resources in legumes

Species	Type of population	Population size (nos)	Mutagen used	Remarks	References
Medicago	Induced	5,000 M1 and 5,000 M_2	EMS (0.2 %)	9–12 alleles/1 kb target sequence	http://www.gl-http.com/products_services/technical_services/genomic_resources_from_glip/functional_genomics
		8,000 M_2		9.8 alleles	Triques et al. (2007)
Pea	Induced	48,000 M_2 and 5,000 M_2	EMS (0.2–0.3 %)	10–40 alleles/1 kb target sequence	-do-
	Natural	400 germplasm lines	–	Identification of hidden allele and association mapping	-do-
	Induced	4,817 lines	93 symbiotic mutants 26 genes involved in nitrogen fixation		Hofer et al. (2009)
Mungbean	Natural	25 lines	–	Single nucleotide polymorphisms (SNPs) and small insertions/deletions (INDELS) in a collection of *Vigna radiata* accessions	Barkley et al. (2008)
L. japonicus	Induced	45,600 M_2	(1 % v/v)	Population comprises to morphological, symbiotic and metabolic mutants of *L. japonicus*. Population was used to identify mutants for sucrose synthesis genes (LjSUS1 to LjSUS4)	Perry et al. (2003); Horst et al. (2007); Stracke et al. (2002)
	Induced	4,904 M_2	EMS		Perry et al. (2009)

Gulberga in a molecular breeding network project funded by Department of Biotechnology, Government of India. In lentil also this approach has been successfully applied with 21% F_1 plants identified at true F_1s and the others as selfed or admixtures (Solanki et al. 2010).

Diversity Analysis Studies

Moleculer markers greatly help in studying the availability and level of genetic diversity among the different gene-pools (Zong et al. 2009; Taunk et al. 2012). Diversity analysis studies in food legumes which have a comparatively narrow genetic base may also help in identifying contrast parents for development of ideal mapping population for a variety of uses. Comprehensive assessment of genetic diversity help identify and rescue the genetic resources at the verge of extinction (Polegri and Negri 2010). The genetic diversity estimates using molecular markers in different crops including pea demonstrated that no gain or reduction of genetic diversity has occurred in last five decades (van de Wouw et al. 2010).

Marker Assisted Breeding

Marker assisted recurrent selection (MARS) and marker-assisted backcrossing (MABC) are the two approaches of marker assisted breeding in legumes as well as other crops. MABC involves introgression of specific trait(s) from a donor parent into the genetic background of a recurrent parent using molecular markers (Hospital 2005). This approach can also be used to generate near-isogenic lines (NILs) or chromosome segment substitution lines (CSSLs) for genomics research, which are populations that are often used for genetic analysis of genes/QTLs and alien gene introgressions (Varshney et al. 2013b).Use of MAS is especially advantageous for traits with low heritability where traditional selection in difficult, expansive, or lacks accuracy or precision (Varshney et al. 2010).

MARS is used to estimate the marker effects from genotyping F_2 or F_3 population and phenotyping F_2 derived F_4 or F_5 progenies, followed by two or three recombinant cycles based on presence of marker alleles for small effect QTLs (Eathington et al. 2007). For MARS, identification of QTL in the population (generally good × good cross) is followed by crossing the lines carrying superior alleles for maximum QTLs to pyramid superior alleles in a single genetic background. The resultant recombinant lines are screened finally in the field to identify the best lines for their multi-locational evaluation and their possible release as a cultivar. The genetic gain achieved in MARS is higher because it captures several genomic regions at a time, and more number of major and minor QTLs (Bernardo and Charcosset 2006).

Knowledge of marker-trait association provides greater insight to the breeders in executing MAS in a better way for development of improved cultivars. The manipulation of the genomic regions having positive additive effects on traits of

interest can lead to maximum potential genetic gain through MAS, particularly for traits having low heritabilities and difficulties in scoring (Kumar et al. 2011). Soybean is the best example where use of markers in breeding programmes has been most successfully demonstrated (Pratap et al. 2012). In past several years, many improved varieties/lines for resistance to different SCN races (Arelli and Young 2009), phytophthora root rot and brown stem rot, insect resistance (Warrington et al. 2008); low linolenic acid content, yield (Concibido et al. 2003), mosaic virus resistance (Shi et al. 2009) have been developed. MAS has also been used successfully in common-bean to develop several lines which are resistant to rust (Stavely 2000; Faleiro et al. 2001), anthracnose (Alzate-Marin et al. 1999) and bean golden yellow mosaic virus (Miklas 2002). In peanut, markers linked with root knot nematode resistance were introgressed into cultivated background via amphidiploids pathway (Simpson et al. 2001). DNA fragment carrying nematode resistance gene was also introgressed selecting a recessive AhFAD2B allele using the linked markers for foreground selection (Chu et al. 2011). This led to development and release of the improved variety "Tiftguard High O/L". Currently, MABC is also being practiced for introgression and pyramiding *Fusarium* wilt and *Ascochyta* blight resistance gene into chickpea in India (Chamarthi et al. 2011; Varshney et al. 2012) by ICRISAT, IIPR and other collaborators in state agricultural universities. In one such project funded by Department of Biotechnology, Government of India, resistance to two races (*foc2* and *foc 4*) independently and pyramiding of resistance to two races (*foc1* and *foc3*) of fusarium wilt and two QTLs for resistance to Acsochyta blight is being undertaken using MABC and MARS and currently, various generations (BC_1F_2 to $BC_3F_{3:4}$) are available for the different crosses. Similarly, for drought tolerance, nine different chickpea varieties have been targeted (see Varshney et al. 2012). Efforts have also been initiated to use MARS in chickpea at ICRISAT, IARI and IIPR.

Gene pyramiding is also a useful approach to achieve multiple and durable resistance (Shi et al. 2009). It has been successfully demonstrated in soybean where genes controlling resistance to CSN have been pyramided (Concibido et al. 2004). Similarly, QTLs/genes controlling tolerance to Phytophthora root rot and resistance to soybean mosaic virus have also been stacked in this crop (Shi et al. 2009; Li et al. 2010).

MAS for two QTLs available on separate linkage groups has been shown to be effective in imparting white mould resistance in common bean (Ender et al. 2008). Similarly, MAS for a major QTL associated with root-rot resistance was found to be effective and it imparted realized gain in plant biomass and vigour traits associated with root-rot complex in snap bean (Navarro et al. 2009). Utilization of MAS has also resulted in development of several improved cultivars in common bean and soybean, mostly in USA (Chamarthi et al. 2011; Pratap et al. 2012). In common bean three genotypes, USPT-ANT-1, ABCP-8 and ABC-Weihing have been released between 2004 and 2006 (Miklas et al. 2003; Mutlu et al. 2008). Similarly, a number of varieties (JTN5503, JTN5303, JTN5109, DS880) have been released in soybean also for resistance to diseases and soybean cyst nematode (Arelli et al. 2006, 2007; Arelli and Young 2009; Smith 2010).

Genome-Wide Selection

Genome-wide selection (GWS) or "genomic selection (GS)" is useful for complex traits that are controlled by many genes/ QTL, each with small effect (Chamarthi et al. 2011). This method predicts genomic estimated breeding values (GEBVs) of progenies, which are calculated for progenies, based on both phenotyping and genotyping data. These GEBVs are then used to select the superior progeny lines for advancement in the breeding cycle (Heffner et al. 2009; Jannink et al. 2010). Doubled haploid (DH) populations are very useful in GWS compared to F_2 populations, when many QTL control a trait (Mayor and Bernardo 2009). GWS can help breeders in reducing the frequency of extensive phenotyping as well as bypass the need of QTL mapping besides reducing the selection cycle, thereby having considerable savings of time. However, there is not much information available on use of GWS in legumes, although recent developments in plant genomics make it feasible to generate genome-wide marker data (using SNPs) to start GWS in breeding programmes. In the coming few years, GWS is expected to be used at least in soybean among the legumes.

Advanced Backcross QTL Analysis and Harnessing Variability from Secondary Gene Pool

Many a times the genes for traits of interest may not be available in cultivated/ primary gene pool of a species and it is necessary to explore the wild species/relative for them. However, owing to linkage drag, their use in conventional breeding programmes still remains restricted. It is now possible to recover the favourable alleles in elite germplasm avoiding associated linkage drag using molecular maps and integrative analysis. In the advanced backcross QTL (AB-QTL) approach, parallel discovery and transfer of desired QTL from an unadapted germplasm into selected breeding lines takes place (Tanksley and Nelson 1996). In AB-QTL, repeated backcrossing is done with the elite parent in wild × cultivated species cross and selection is imposed in advanced backcrossed (BC_2F_2 or BC_2F_3) populations. This approach reduces linkage drag as well generates phenotyping and genotyping data. The advanced backcross populations are simultaneously used to identify desirable genes/QTL through QTL analysis. Once favourable QTL alleles are identified, marker assisted selection in a few generations (3–4) can lead to development of near isogenic lines (NILs) which can be used for development of a variety. This approach has been successfully used in soybean and commonbean (Blair et al. 2003; Chaky et al. 2003). Foncéka et al. (2009) reported a successful effort for genome wide segment introgressions from a synthetic amphidiploids (*A. duranensis*×*A. ipaensis*) to a cultivated variety (Fluer 11) using molecular markers. The backcross (BC_1F_1 and BC_2F_1) lines carrying the wild genome segments with maximum recurrent parent genomic regions provided optimal distribution of the synthetic genome introgressions.

In another approach, introgression libraries are constructed which are made up of several introgression lines (ILs). The ILs are developed by repeated backcrossing of F_1s between wild × cultivated lines. This leads to distribution of donor (wild species) genome into the entire genome of ILs and consequently their expression in the phenotype. Such libraries have been reported to be developed in soybean using wild soybean species (*G. soja*) (Concibido et al. 2003) and groundnut from synthetic tetraploids (Foncéka et al. 2009).

Conclusions and Perspectives

In the past decade, proactive and coordinated efforts of the international legume community have ensured a significant progress in the development of genomic resources of food legumes which have led to a better understanding of their genome structure. These have also offered new possibilities for genetic improvement of not only grain legumes but also several other species, especially those where their development is costly. While the cost effective, polymorphic and reproducible markers such as SSRs, SNPs, etc. can be used by breeders in development of improved cultivars through marker assisted breeding employing MAS, MARS and MABC, high throughput sequencing can accelerate the development of new molecular markers. The marker-trait association will enable biotechnologists to more rapidly and precisely manipulate target genes underlying key agronomic traits, especially a series of abiotic and biotic stresses limiting crop productivity. This will be especially useful in developing such genotypes which suit the marginal environments of food legume growing areas of the world. Increased focus is required on development of organized genome resources including physical maps and functional genomic tools, TILLING populations, and microarray chips, which will facilitate the isolation of genes for resistance/tolerance to biotic and abiotic stresses. Ultimately, the availability of high-throughput and cost-effective genotyping platforms, combined with automation in phenotyping methodologies, will increase the uptake of genomic tools into breeding programs, and thus usher an era of genomics-enabled molecular breeding in legumes.

References

Allen RS, Millgate AM, Chitty JA, Thistleton J, Miller JAC, Fist AJ, Gerlach WL, Larkin PJ (2004) RNAi-mediated replacement of morphine with the non-narcotic alkaloid reticuline in opium poppy. Nat Biotechnol 22:1559–1566

Alzate-Marin AL, Menarim H, Arruda MCC, Chagas JM, Barros EG, Moreira MA (1999) Backcross assisted by RAPD markers for introgression of Co-4 and Co-6 anthracnose resistant genes in common bean cultivars. Ann Rep Bean Improv Coop 42:15–16

Arelli PR, Young LD (2009) Jtn-5109 soybean germplasm resistant to nematode population infecting cv. Hartwig. Agronomy Society of America, The Abstracts 268–18: 133 (http://a-c-s.confex.com/crops/2009am/webprogram/Paper51979.html)

Arelli PR, Young LD, Mengistu A (2006) Registration of high yielding and multiple disease resistant soybean germplasm Jtn-5503. Crop Sci 46:2723–2724

Arelli PR, Pantalone VR, Allen FL, Mengistu A (2007) Registration of soybean germplasm Jtn-5303. J Plant Reg 1:69–70

Ashraf N, Ghai D, Barman P, Basu S, Gangisetty N, Mandal R, Chakraborty N, Datta A, Chakraborty S (2009) Comparative Analysis of genotype dependent expressed sequenced tag and stress responsive transcriptome of chick pea will illustrate predicted and unpredicted genes and novel regulators of plant immunity. BMC Genomics 10:415

Aziz N, Paiva NL, May GD, Dixon RA (2005) Transcriptome analysis of alfalfa glandular trichomes. Planta 221:28–38

Baier MC, Barsch A, Kuster H, Hohnjec N (2007) Antisense repression of the *Medicago truncatula* nodule-enhanced sucrose synthase leads to a handicapped nitrogen fixation mirrored by specific alterations in the symbiotic transcriptome and metabolome. Plant Physiol 145:1600–1618

Barkley NA, Wang ML, Gillaspie AG, Dean RE, Pederson GA, Jenkins TM (2008) Discovering and verifying DNA polymorphisms in a mung bean [*V. radiata* (L.) R. Wilczek] collection by EcoTILLING and sequencing. BMC Res Notes 1:28–35

Benedito VA, Torres-Jerez I, Murray JD, Andriankaja A, Allen S, Kakar K, Wandrey M, Verdier J, Zuber H, Ott T, Sandra M, Andreas N, Tancred F, Georg W, Ji H, Xinbin D, Patrick ZX, Yuhong T, Michael KU (2008) A gene expression atlas of the model legume *Medicago truncatula*. Plant J 55:504–513

Benlloch R, D'Erfurth I, Ferrandiz C, Cosson V, Beltran JP, Canas LA, Kondorosi A, Madueno F, Ratet P (2006) Isolation of mtpim proves *Tnt1* a useful reverse genetics tool in *Medicago truncatula* and uncovers new aspects of *AP1* functions in legumes. Plant Physiol 142:972–983

Bennett MD, Leitch IJ (2012) Plant DNA C-values database. Available at: http://www.kew.org/cvalues/. Accessed 2 Apr 2013

Bernardo R, Charcosset A (2006) Usefulness of gene information in marker-assisted recurrent selection: a simulation appraisal. Crop Sci 46:614–621

Blair MW, Pedraza F, Buendia HF, Gatian-Soils E, Beebe SE, Gepts P, Tohme J (2003) Development of a genome-wide anchors microsatellite map for common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 107:1362–1374

Bohra A, Dubey A, Saxena RK, Penmetsa RV, Poornima KN, Kumar N (2011) Analysis of BAC-end sequences (BESs) and development of BES-SSR markers for genetic mapping and hybrid purity assessment in pigeonpea (*Cajanus* spp.). BMC Plant Biol 11:56

Bohra A, Saxena RK, Gnanesh BN, Saxena K, Byregowda M, Rathore A, Kavi Kishor PB, Cook DR, Varshney RK (2012) An intra-specific consensus genetic map of pigeonpea [*Cajanus cajan* (L.) Millspaugh] derived from six mapping populations. Theor Appl Genet 125: 1325–1338. http://link.springer.com/journal/122

Briñez B, Blair MW, Kilian A, Carbonell SAM, Chiorato AF, Rubiano LB (2011) A whole genome DArT assay to assess germplasm collection diversity in common beans. Mol Breed. doi:10.1007/s11032-011-9609-3

Buhariwalla HK, Jayashree B, Eshwar K, Crouch JH (2005) Development of ESTs from chickpea roots and their use in diversity analysis of the *Cicer* genus. BMC Plant Biol 5:16–29

Buitink J, Leger JJ, Guisle I, Vu BL, Wuilleme S, Lamirault G, Le Bars A, Le Meur N, Becker A, Kuster H, Leprince O (2006) Transcriptome profiling uncovers metabolic and regulatory processes occurring during the transition from desiccation-sensitive to desiccation-tolerant stages in *Medicago truncatula* seeds. Plant J 47:735–750

Chaky JM, Specht JE, Cregan PB (2003) Advanced backcross QTL analysis in a mating between *Glycine max* and *Glycine soja*. Plant Anim Genome Abstr 545

Chamarthi SK, Kumar A, Vuong TD, Blair MW, Gaur PM, Nguyen HT (2011) Trait mapping and molecular breeding. In: Pratap A, Kumar J (eds) Biology and breeding of food legumes. CABI, Oxfordshire, UK, pp 297–313

Chandran D, Sharopova N, Ivashuta S, Gantt JS, Vandenbosch KA, Samac DA (2008) Transcriptome profiling identified novel genes associated with aluminum toxicity, resistance and tolerance in *Medicago truncatula*. Planta 228:151–166

Choudhary S, Gaur R, Gupta S, Bhatia S (2012) EST derived genetic molecular marker: development and utilization for generating an advanced transcript map of chick pea. Theor Appl Genet 124:1449–1462

Chu Y, Wu CL, Holbrook CC, Tillman BL, Person G, Ozias-Akins P (2011) Marker-assisted selection to pyramid nematode resistance and the high oleic trait in peanut. Plant Genome 4:110–117

Concibido VC, Vallee BL, Mclaird P, Pineda N, Meyer J, Hummel L, Yang J, Wu K, Delannay X (2003) Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. Theor Appl Genet 106:575–582

Concibido VC, Diers BW, Arelli PR (2004) A decade of QTL mapping for cyst nematode resistance in soybean. Crop Sci 44:1121–1131

Constantin GD, Krath BN, MacFarlane SA, Nicolaisen M, Johansen IE, Lund OS (2004) Virus-induced gene silencing as a tool for functional genomics in a legume species. Plant J 40:622–631

Coram TE, Pang ECK (2005a) Isolation and analysis of candidate Ascochyta blight defence genes in chickpea. Part I. Generation and analysis of an expressed sequence tag (EST) library. Physiol Mol Plant Pathol 66:192–200

Coram TE, Pang ECK (2005b) Isolation and analysis of candidate Ascochyta blight defence genes in chickpea. Part II. Microarray expression analysis of putative defence-related ESTs. Physiol Mol Plant Pathol 66:201–210

Córdoba JM, Chavarro C, Schlueter JA, Jackson SA, Blair MW (2010) Integration of physical and genetic maps of common bean through BAC-derived microsatellite markers. BMC Genomics 11:436

Cortés AJ, Chavarro MC, Blair MW (2011) SNP marker diversity in common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 123:827–845

Coyne CJ, McClendon MT, Walling JG, Timmerman-Vaughan GM, Murray S, Meksem K, Lightfoot DA, Shultz JL, Keller KE, Martin RR, Inglis DA, Rajesh PN, McPhee KE, Weeden NF, Grusak MA, Li CM, Storlie EW (2007) Construction and characterization of two bacterial artificial chromosome libraries of pea (*Pisum sativum* L.) for the isolation of economically important genes. Genome 50:871–875

Cregan PB, Mudge J, Fickus EW, Danesh D, Denny R, Young ND (1996) Two simple sequence repeat markers to select for soybean cyst nematode resistance conditioned by the *rhg1* locus. Theor Appl Genet 99:811–818

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG et al (1999) An integrated genetic linkage map of the soybean genome. Crop Sci 39:1464–1490

D'Erfurth I, Cosson V, Eschstruth A, Lucas H, Kondorosi A, Ratet P (2003) Efficient transposition of the *Tnt1* tobacco retrotransposon in the model legume *Medicago truncatula*. Plant J 34:95–106

Dubey A, Farmer A, Schlueter J, Cannon SB, Abernathy B, Tuteja R (2011) Defining the transcriptome assembly and its use for genome dynamics and transcriptome profiling studies in pigeonpea (*Cajanus cajan* L.). DNA Res 18:153–164

Dutta S, Kumawat G, Singh BP, Gupta DK, Singh S, Dogra V (2011) Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea (*Cajanus cajan* (L.) Millspaugh). BMC Plant Biol 11:17

Eathington SR, Crosbie TM, Edwards MD, Reiter RS, Bull JK (2007) Molecular markers in a commercial breeding program. Crop Sci 47:S154–S163

Ender M, Terpstra K, Kelly JD (2008) Marker-assisted selection for white mold resistance in common bean. Mol Breed 21:149–157

Endre G, Kereszt A, Kevei Z, Mihacea S, Kalo P, Kiss GB (2002) A receptor kinase gene regulating symbiotic nodule development. Nature 417:962–966

Faleiro FG, Ragagnin VA, Carvalho GA, Paula TJ Jr, Moreira MA, Barros EG (2001) Development of common bean lines resistant to rust and anthracnose by molecular marker-assisted backcrossing. Annu Rep Bean Improvement Coop 44:1130–1133

Firnhaber C, Puhler A, Kuster H (2005) EST sequencing and time course microarray hybridizations identify more than 700 *Medicago truncatula* genes with developmental expression regulation in flowers and pods. Planta 222:269–283

Foncéka D, Hodo-Abalo T, Rivallan R, Faye I, Sall MN, Ndoye O (2009) Genetic mapping of wild introgressions into cultivated peanut: a way toward enlarging the genetic basis of a recent allotetraploid. BMC Plant Biol 9:103

Galeano CH, Fernandez AC, Franco-Herrera N, Cichy KA, McClean PE (2011) Saturation of an intra-gene pool linkage map: towards a unified consensus linkage map for fine mapping and synteny analysis in common bean. PLoS One 6(12):e28135. doi:10.1371/journal.pone.0028135

Garg R, Patel RK, Jhanwar S, Priya P, Bhattacharjee A, Yadav G (2011a) Gene discovery and tissue-specific transcriptome analysis in Chickpea with massively parallel pyrosequencing and web resource development. Plant Physiol 156:1661–1678

Garg R, Patel RK, Tyagi AK, Jain M (2011b) *De Novo* assembly of chickpea transcriptome using short reads for gene discovery and marker identification. DNA Res 18:53–63

Gaur R, Sethy NK, Choudhary S, Shokeen B, Gupta V, Bhatia S (2011) Advancing the STMS genomic resources for defining new locations on the intraspecific genetic linkage map of chickpea (*Cicer arietinum* L.). BMC Genomics 12:117

Gaur PM, Jukanti AK, Varshney RK (2012) Impact of genomic technologies on chickpea breeding strategies. Agronomy 2:199–221

Gepts P, Beavis WD, Brummer CE, Shoemaker RC, Stalker TH, Weeden NF, Young ND (2005) Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through genomics conference. Plant Physiol 137:1228–1235

Gepts P, Aragão F, de Barros E, Blair MW, Brondani R, Broughton W, Galasso I, Hernández G, Kami J, Lariguet P, McClean P, Melotto M, Miklas P, Pauls P, Pedrosa-Harand A, Porch T, Sánchez F, Sparvoli F, Yu K (2008) Genomics of *Phaseolus* beans, a major source of dietary protein and micronutrients in the tropics. In: Moore P, Ming R (eds) Genomics of tropical crop plants. Springer, New York, NY, pp 113–143

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163–185

Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. Crop Sci 49:1–12

Hiremath PJ, Farmer A, Cannon SB, Woodward J, Kudapa H, Tuteja R (2011) Large-scale transcriptome analysis in chickpea (*Cicer arietinum* L.), an orphan legume crop of the semi-arid tropics of Asia and Africa. Plant Biotechnol J 9:922–931

Hiremath PJ, Kumar A, Penmetsa RV, Farmer A, Schlueter JA, Chamarthi SK (2012) Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. Plant Biotechnol J. doi:10.1111/j.1467-7652.2012.00710.x

Hofer J, Turner L, Moreau C, Ambrose M, Isaac P, Butcher S, Weller J, Dupin A, Dalmais M, Le Signor C, BendahmaneA EN (2009) *Tendril-less* regulates tendril formation in pea leaves. Plant Cell 21:420–428

Horst I, Welham T, Kelly S, Kaneko T, Sato S, Tabata S, Parniske M, Wang TL (2007) TILLING mutants of *Lotus japonicus* reveal that nitrogen assimilation and fixation can occur in the absence of nodule-enhanced sucrose synthase. Plant Physiol 144:806–820

Hospital F (2005) Selection in backcross programmes. Phil Trans R Soc B 360:1503–1511

Hyten DL, Song O, Fickus EW, Quigley CV, Lim JS, Choi IY, Hwang EY, Pastor Corrales M, Cregan PB (2010) High-throughput SNP discovery and assay development in common bean. BMC Genomics 11:475

Isemura T, Kaga A, Tabata S, Somta P, Srinives P et al (2012) Construction of a genetic linkage map and genetic analysis of domestication related traits in mungbean (*Vigna radiata*). PLoS One 7(8):e41304. doi:10.1371/journal.pone.0041304

Jackson SA, Rokhsar D, Stacey G, Shoemaker RC, Schmutz J, Grimwood J (2006) Toward a reference sequence of the soybean genome: a multiagency effort. Crop Sci 46:555–561

Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genom Prot 9:166–177

Jayashree B, Buhariwalla HK, Shinde S, Crouch JH (2005) A legume genomics resource: the chickpea root expressed sequence tag database. Electron J Biotechnol 8:128–133

Jones KM, Sharopova N, Lohar DP, Zhang JQ, VandenBosch KA, Walker GC (2008) Differential response of the plant *Medicago truncatula* to its symbiont *Sinorhizobium meliloti* or an exopolysaccharide-deficient mutant. Proc Natl Acad Sci U S A 105:704–709

Kaczorowski KA, Kim KS, Diers BW, Hudson ME (2008) Microarray-based genetic mapping using soybean near-isogenic lines and generation of SNP markers in the *Rag1* aphid-resistance interval. Plant Genome 1:89–98

Kassa MT, Penmetsa RV, Carrasquilla-Garcia N, Sarma BK, Datta S, Upadhyaya HD (2012) Genetic patterns of domestication in pigeonpea (*Cajanus cajan* (L.) Millsp.) and wild *Cajanus* relatives. PLoS One 7:e39563

Kudapa H, Bharti AK, Cannon SB, Farmer AD, Mulaosmanovic B, Kramer R (2012) A comprehensive transcriptome assembly of pigeonpea (*Cajanus cajan* L.) using Sanger and second-generation sequencing platforms. Mol Plant 1–9; doi:10.1093/mp/ssr111

Kuester H, Hohnjec N, Krajinski F, El Yahyaoui F, Manthey K, Gouzy J, Dondrup M, Meyer F, Kalinowski J, Brechenmacher L, van Tuinen D, Gianinazzi-Pearson V, Pühler A, Gamas P, Becker A (2004) Construction and validation of cDNA-based Mt6k-RIT macroarray and microarray to explore root endosymbioses in the model legume *Medicago truncatula*. J Biotechnol 108:95–113

Kuhar K, Gupta VK, Kansal R, Gupta VK (2012) Isolation and *in silico* characterization of cDNA encoding cyclophilin from etiolated *Vigna mungo* seedlings. Braz J Plant Physiol 24:69–73

Kumar J, Choudhary A, Solanki RK, Pratap A (2011) Towards marker-assisted selection in pulses—a review. Plant Breed 130:297–313

Lee Y, Tsai J, Sunkara S, Karamycheva S, Pertea G, Sultana R, AntonescuV CA, Cheung F, Quackenbush J (2005) The TIGR gene indices: clustering and assembling EST and known genes and integration witheukaryotic genomes. Nucleic Acids Res 33:D71–D74

Lewers K, Heinz R, Beard H, Marek L, Matthews B (2002) A physical map of a gene- dense region in soybean linkage group A2 near the black seed coat and *Rhg4* loci. Theor Appl Genet 104:254–260

Li X, Han Y, Teng W, Zhang S, Yu K, Poysa V, Anderson T, Ding J, Li W (2010) Pyramided QTL underlying tolerance to phytophthora root rot in mega-environments from soybean cultivars Conrad and Hefeng 25. Theor Appl Genet. doi:10.1007/S00122-010-1337-2

Liu N, Shan Y, Wang FP, Xu G, Peng KM, Li XH, Zhang Q (2001) Identification of an 85-kb DNA fragment containing *pms 1*, a locus for photoperiod-sensitive genic male-sterility in rice. Mol Genet Genomics 266:271–275

Liu J, Blaylock LA, Endre G, Cho J, Town CD, VandenBosch KA, Harrison MJ (2003) Transcript profiling coupled with spatial expression analyses reveals genes involved in distinct developmental stages of an arbuscular mycorrhizal symbiosis. Plant Cell 15:2106–2123

Lohar DP, Sharopova N, Endre G, Penuela S, Samac D, Town C, Silverstein KAT, VandenBosch KA (2006) Transcript analysis of early nodulation events in *Medicago truncatula*. Plant Physiol 140:221–234

Lucas MR, Diop N, Wanamaker S, Ehlers JD, Roberts PA, Close TJ (2011) Cowpea–soybean synteny clarified throughan improved genetic map. Plant Genome 4:218–225

Mallikarjuna N, Senapathy S, Jadhav DR, Saxena KB, Sharma HC, Upadhyaya HD (2011) Progress in the utilization of *Cajanus platycarpus* in pigeonpea improvement. Plant Breed 130:507–514

Mantri NL, Ford R, Coram TE, Pang ECK (2007) Transcriptional profiling of chickpea genes differentially regulated in response to high-salinity, cold and drought. BMC Genomics 8:303–316

Mantri NL, Coram TE, Ford R, Pang ECK (2010) Evidence of unique and shared responses to major biotic and abiotic stresses in chickpea. Environ Exp Bot 69:286–292

Mayor PJ, Bernardo R (2009) Genomewide selection and marker-assisted recurrent selection in doubled haploid versus F_2 populations. Crop Sci 49:1719–1725

Meyers BC, Galbraith DW, Nelson T, Agrawal V (2004) Methods for transcriptional profiling in plants: be fruitful and replicate. Plant Physiol 135:637–652

Miklas PN (2002) Marker assisted selection for disease resistance in common beans. Ann Rep Bean Improv Coop 45:1–3

Miklas PN, Kelly JD, Singh S (2003) Registration of anthracnose resistant pinto bean germplasm line USPT-ANT-1. Crop Sci 43:1889–1890

Molina C, Rotter B, Horres R, Udupa SM, Besser B, Bellarmino L, Baum M, Matsumura H, Terauchi R, Kahl G, Winter P (2008) SuperSAGE: the drought stress-responsive transcriptome of chickpea roots. BMC Genomics 9:553–580

Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S, Pottorff M, Hearne S, Cisse N, Fatokun C, Ehlers JD, Roberts PA, Close TJ (2009) A consensus genetic map of cowpea [*Vigna unguiculata* (L) Walp.] and synteny based on EST-derived SNPs. Proc Natl Acad Sci 106:18159–18164

Mustafa BM, Coram TE, Pang ECK, Taylor PWJ, Ford R (2009) A cDNA microarray approach to decipher Ascochyta blight resistance in lentil. Austr Plant Pathol 38:617–631

Mutlu N, Urrea CA, Miklas PN, Steadman JR, Pastor Corrales MA, Lindgren DT, Reiser J, Vidaver AK, Coyne DP (2008) Registration of common bacterial blight, rust and bean common mosaic resistant great northern bean germplasm line Abc-weighing. J Plant Reg 2:120–124

Narasimhamoorthy B, Bouton JH, Olsen KM, Sledge MK (2007) Quantitative trait loci and candidate gene mapping of aluminum tolerance in diploid alfalfa. Theor Appl Genet 114:901–913

Navarro FM, Sass ME, Nienhuis J (2009) Marker-facilitated selection for a major QTL associated with root rot resistance in snap bean (*Phaseolus Vulgaris* L.). Crop Sci 49:850–856

Nayak SN, Zhu H, Varghese N, Datta S, Choi HK, Horres R (2010) Integration of novel SSR and gene-based SNP marker loci in the chickpea genetic map and establishment of new anchor points with *Medicago truncatula* genome. Theor Appl Genet 120:1415–1441

Nunberg A, Bedell JA, Budiman MA, Citek RW, Clifton SW, Fulton L, Pape D, Cai Z, Joshi T, Nguyen H, Xu D, Stacey G (2006) Survey sequencing of soybean elucidates the genome structure, composition and identifies novel repeats. Funct Plant Biol 33:765–773

Nunes AC, Vianna GR, Cuneo F, Amaya-Farfan J, deCap deville G, Rech EL, Aragao FJ (2006) RNAi-mediated silencing of the myo-inositol-1-phosphate synthase gene (GmMIPS1) in transgenic soybean inhibited seed development and reduced phytate content. Planta 224:125–132

O'Rourke IJA, Charlson DV, Gonzalez DO, Vodkin LO, Graham MA, Cianzio SR, Grusak MA, Shoemaker RC (2007) Microarray analysis of iron deficiency chlorosis in near-isogenic soybean lines. BMC Genomics 8:476–488

Pandey MK, Monyo E, Ozias-Akins P, Liang X, Guimarães P, Nigam SN et al (2012) Advances in *Arachis* genomics for peanut improvement. Biotechnol Adv 30:639–651

Perry JA, Wang TL, Welham TJ, Gardner S, Pike JM, Yoshida S, Parniske M (2003) A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. Plant Physiol 131:866–871

Perry J, Brachmann A, Welham T, Binder A, Charpentier M, Groth M, Haage K, Markmann K, Wang TL, Parniske M (2009) TILLING in *Lotus japonicas* identified large allelic series for symbiosis genesand revealed a bias in functionally defective ethyl methane sulfonate alleles toward glycine replacements. Plant Physiol 151:1281–1291

Polegri L, Negri V (2010) Molecular markers for promoting agro-biodiversity conservation: a case study from Italy. How cowpea landraces were saved from extinction. Genet Res Crop Evol 57:867–880

Pratap A, Gupta SK, Kumar J, Solanki RK (2012) Soybean. In: Gupta SK (ed) Technological innovations in major world oil crops, Vol I. Breeding. Springer Science + Business Media, New York, NY, pp 293–321

Rajesh PN, Coyne C, Meksem K, DerSharma K, Gupta V, Muehlbauer FJ (2004) Construction of a *Hind*III bacterial artificial chromosome library and its use in identification of clones associated with disease resistance in chickpea. Theor Appl Genet 108:663–669

Raju NL, Gnanesh BN, Pazhamala L, Jayashree B, Pande S, Hiremath PJ, Byregowda M et al (2010) The first set of EST resource for gene discovery and marker development in pigeonpea (*Cajanus cajan* L.). BMC Plant Biol 10:45

Reddy MS, Dinkins RD, Collins GB (2003) Gene silencing in transgenic soybean plants transformed *via* particle bombardment. Plant Cell Rep 21:676–683

Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K, Fujishiro T, Katoh M, Kohara M, Kishida Y, Minami C, Nakayama S, Nakazaki N, Shimizu Y, Shinpo S, Takahashi C, Wada T, Yamada M, Ohmido N, Hayashi M, Fukui K, Baba T, Nakamichi T, Mori H, Tabata S (2008) Genome structure of the legume, *Lotus japonicus*. DNA Res 15:227–239

Saxena KB, Nadarajan N (2010) Prospects of pigeonpea hybrids in Indian agriculture. Electron J Plant Breed 1:1107–1117

Schlueter JA, Goicoechea JL, Collura K, Gill N, Lin JY, Yu Y, Kudrna D, Zuccolo A, Vallejos CE, Muñoz-Torres M, Blair MW, Tohme J, Tomkins J, McClean P, Wing RA, Jackson SA (2008) BAC-end sequence analysis and a draft physical map of the common bean (*Phaseolus vulgaris* L.) genome. Trop Plant Biol 1:40–48

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Shi A, Chen P, Li D, Zheng C, Zhang B, Hou A (2009) Pyramiding multiple genes for resistance to soybean mosaic virus in soybean using molecular markers. Mol Breed 23:113–124

Shoemaker RC, Schlueter JA, Cregan P, Vodkin L (2003) The status of soybean genomics and its role in the development of soybean biotechnologies. Agric Bio Forum 6:4–7

Simpson CE, Krapovickas A, Valls JFM (2001) History of *Arachis* including evidence of *A. hypogaea* L. progenitors. Peanut Sci 28:78–79

Singh NK, Gupta DK, Jayaswal PK, Mahato AK, Dutta S (2012) The first draft of the pigeon pea genome sequence. J Plant Biochem Biotechnol. doi:10.1007/s13562-011-0088-8

Smith J (2010) USDA, ARS, National Genetic Resources Program. Germplasm Resources Information Network – (GRIN) [Online Database]. National Germplasm Resources Laboratory, Beltsville, MD. http://www.ars-grin.gov/cgi-bin/npgs/html/index.pl. Accessed 17 May 2013

Solanki RK, Singh S, Kumar J (2010) Molecular marker-assisted testing of hybridity of F_1 plants in lentil. J Food Leg 23:21–24

Stavely JR (2000) Pyramiding rust and viral resistance genes using traditional and marker techniques in common bean. Ann Rep Bean Improv Coop 43:1–3

Stracke S, Kistner C, Yoshida S, Mulder L, Sato S, Kaneko T, Tabata S, Sandal N, Stougaard J, Szczyglowski K, Parniske M (2002) A plant receptor-like kinase required for both bacterial and fungal symbiosis. Nature 417:959–962

Subramanian S, Graham MY, Yu O, Graham TL (2005) RNA interference of soybean isoflavone synthase genes leads to silencing in tissues distal to the transformation site and to enhanced susceptibility to *Phytophthora sojae*. Plant Physiol 137:1345–1353

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92:191–203

Tao Q, Chang YL, Wang J, Chen H, Islam-Faridi MN, Scheuring C, Wang B, Stelly DM, Zhang HB (2001) Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis. Genetics 158:1711–1724

Taunk J, Yadav NR, Yadav RC, Kumar R (2012) Genetic diversity among green gram (*Vigna radiata* L. Wilczek) genotypes varying in micronutrient content using RAPD. Ind J Bitechnol 11:48–53

Thudi M, Bohra A, Nayak SN, Varghese N, Shah TM, Penmetsa RV (2011) Novel SSR markers from BAC-end sequences, DArT arrays and a comprehensive genetic map with 1,291 marker loci for chickpea (*Cicer arietinum* L.). PLoS One 6:e27275

Timko MP, Rushton PJ, Laudeman TW, Bokowiec MT, Chipumuro E, Cheung F, Town CD, Chen X (2008) Sequencing and analysis of the gene-rich space of cowpea. BMC Genomics 9:103

Triques K, Sturbois B, Gallais S, Dalmais M, Chauvin S, Clepet C, Aubourg S, Rameau C, Caboche M, Bendahmane A (2007) Characterization of *Arabidopsis thaliana* mismatch specific endonucleases: application to mutation discovery by TILLING in pea. Plant J 51: 1116–1125

Van de Wouw M, Van Hinten T, Kik C, Van Treuren R, Visser B (2010) Genetic diversity trends in twentieth century crop cultivars: a meta analysis. Theor Appl Genet 120:1241–1252

Vance CP, Graham PH, Allan DL (2000) Biological nitrogen fixation phosphorus: a critical future need. In: Pedrosa FO, Hungria M, Yates MG, Newton WE (eds) Nitrogen fixation: from molecules to crop productivity. Kluwer, Dordrecht, pp 506–514

Vanden Bosch K, Stacey G (2003) Summaries of legume genomics projects from around the globe. Community resources for crops and models. Plant Physiol 131:840–865

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55

Varshney RK, Bertioli DJ, Moretzsohn MC, Vadez V, Krishnamurthy L, Aruna R, Nigam SN, Moss BJ, Seetha K, Ravi K, He G, Knapp SJ, Hoisington DA (2009a) The first SSR-based genetic linkage map for cultivated groundnut (*Arachis hypogaea* L.). Theor Appl Genet 118:729–739

Varshney RK, Close TJ, Singh NK, Hoisington DA, Cook DR (2009b) Orphan legume crops enter the genomics era! Curr Opin Plant Biol 12:1–9

Varshney RK, Thudi M, May GD, Jakson SA (2010) Legume genomics and breeding. Plant Breed Rev 33:257–304

Varshney RK, Kudapa H, Roorkiwal M, Thudi M, Pandey MK, Saxena RK et al (2012) Advances in genomics research and molecular breeding applications in SAT legume crops by using next generation sequencing and high-throughput genotyping technologies. J Biosci 37:811–820

Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'an B, Millan T, Zhang X, Ramsay LD, Iwata A, Wang Y, Nelson W, Farmer AD, Gaur PM, Soderlund C, Penmetsa RV, Xu C, Bharti AK, He W, Winter P, Zhao S, Hane JK, Carrasquilla-Garcia N, Condie JA, Upadhyaya HD, Lu MC, Thudi M, Gowda CLL, Singh NP, Lichtenzveig J, Gali KK, Rubio J, Nadarajan N, Dolezel J, Bansal KC, Xu X, Edwards D, Zhang G, Kahl G, Gil J, Singh KB, Datta SK, Jackson S, Wang J, Cook DR (2013a) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. Nat Biotechnol. doi:10.1038/nbt.2491

Varshney RK, Mohan SM, Gaur PM, Gangarao NVPR, Pandey MK, Bohra A, Sawargaonkar SL, Gorantla A, Kimurto PK, Janila P, Saxena KB, Fikre A, Sharma M, Rathore A, Pratap A, Tripathi S, Datta S, Chaturvedi SK, Mallikarjuna N, Anuradha G, Babbar A, Choudhary AK, Mhase MB, Bharadwaj Ch, Mannur DM, Harer PN, Guo B, Liang X, Nadarajan N, Gowda CLL (2013b) Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. Biotechnol Adv (http://dx.doi.org/10.1016/j.biotechadv.2013.01.001)

Verdier J, Kakar K, Gallardo K, Le Signor C, Aubert G, Schlereth A, Town CD, Udvardi MK, Thompson RD (2008) Gene expression profiling of *Medicago truncatula* transcription factors identifies putative regulators of grain legume seed filling. Plant Mol Biol 67:567–580

Vodkin LO, Khanna A, Shealy R, Clough SJ, Gonzalez DO, Philip R, Zabala G, Thibaud-Nissen F, Sidarous M, Stromvik MV, Shoop E, Schmidt C, Retzel E, Erpelding J, Shoemaker RC, Rodriguez-Huete AM, Polacco JC, Coryell V, Keim P, Gong G, Liu L, Pardinas J, Schweitzer P (2004) Microarrays for global expression constructed with a low redundancy set of 27,500 sequenced cDNAs representing an array of developmental stages and physiological conditions of the soybean plant. BMC Genomics 5:73–90

Warrington CV, Zhu S, Parrot WA, All JN, Boerma HR (2008) Seed yield of near isogenic soybean lines introgressed with quantitative trait loci conditioning resistance to corn earworm (Lepidopter: Noctuidae) and soybean looper (Lepidoptera: Noctuidae) from PI 229358. J Econ Entomol 101:1471–1477

Wu C, Sun S, Nimmakayala P, Santos FA, MeksemK SR, Ding K, Lightfoot DA, Zhang HB (2004) A Bac- and Bibac-based physical map of the soybean genome. Genome Res 14:319–326

Xu M, Song J, Cheng Z, Jiang J, Korban SS (2001) A bacterial artificial chromosome (BAC) library of *Malus floribunda* 821 and contig construction for positional cloning of the apple scab resistance gene *Vf*. Genome 44:1104–1113

Yahyaoui FE, Kuster H, Amor BB, Hohnjec N, Puhler A, Becker A, Gouzy J, Vernie T, Gough C, Niebel A, Godiard L, Gamas P (2004) Expression profiling in *Medicago truncatula* identifies

more than 750 genes differentially expressed during nodulation, including many potential regulators of the symbioticprogram. Plant Physiol 136:3159–3176

Yang SY, Saxena RK, Kulwal PA, Ash GJ, Dubey A, Harper DI (2011) The first genetic map of pigeonpea based on diversity arrays technology (DArT) markers. J Genet 90:103–109

Young ND, Udvardi M (2009) Translating *Medicago truncatula* genomics to crop legumes. Curr Opin Plant Biol 12:193–201

Young ND, Cannon SB, Sato S, Kim D, Cook DR, Town CD, Roe BA, Tabata S (2005) Sequencing the gene spaces of *Medicago truncatula* and *Lotus japonicus*. Plant Physiol 137:1174–1181

Yu K (2012) Bacterial artificial chromosome libraries of pulse crops: Characteristics and applications. J Biomedicine Biotechnol. Article ID 493186 doi:10.1155/2012/493186

Zhang X, Scheuring CF, Zhang M, Dong JJ, Zhang Y, HuangJJ LMK, Abbo S, Sherman A, Shtienberg D, Chen W, Muehlbauer F, Zhang HB (2010) A BAC/BIBAC-based physical map of chickpea, *Cicer arietinum* L. BMC Genomics 11:501

Zong X, Redden R, Liu Q, Wang S, Guan J, Liu J, Xu Y, Liu X, Gu J, Yan L, Ades P, Ford R (2009) Analysis of a diverse global *Pisum* sp. collection and comparison to a Chinese local *P. sativum* collection with microsatellite markers. Theor Appl Genet 118:193–204

Chapter 12
Bioinformatics for Legume Genomics Research

Vinay Kumar Singh, A.K. Singh, Arvind M. Kayastha, and B.D. Singh

Abstract Enormous legume genome sequence data are becoming available at a rapid rate through the Next-Gen Sequencing platforms. One of the biggest problems relates to management and analysis of the huge data derived from whole genome sequencing projects. To resolve this problem, researchers index their data in major biological depository systems and availability of algorithms, tools, softwares and databases and provide opportunities for analysis, annotation, and visualization of sequence data at the computational level. Different types of tools and softwares are available for the interpretation of genomes, proteomes and genes. Now researchers are using various *in-silico* techniques in *Bio-omics* (genomics, proteomics, metabolomics and transcriptomics) era for management, planning and prediction of data in cost effective and less time consuming manner. *Bio-omics* plays an important role in comparative, structural and functional biology at computational level and will play major role in different biological investigations. Identification of signal transduction pathway-associated members and gene family members will help in functional elucidation and relationship among them. In this context identification of potential candidate genes will provide an opportunity to researchers for improvement and nutritional quality enhancement of crop genomes. Based on genome blue-prints (plants, animals, fungus, microbes) one can develop potential applications to understand systems biology of legumes in fullness.

V.K. Singh, M.Sc. (✉)
Faculty of Sciences, Centre for Bioinformatics, School of Biotechnology,
Banaras Hindu University, Varanasi, Uttar Pradesh 221005, India
e-mail: vinaysingh@bhu.ac.in

A.K. Singh, Ph.D.
Department of Genetics and Plant Breeding, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi, Uttar Pradesh, India

A.M. Kayastha, Ph.D. • B.D. Singh, Ph.D.
Faculty of Sciences, School of Biotechnology, Banaras Hindu University,
Varanasi, Uttar Pradesh, India

S. Gupta et al. (eds.), *Legumes in the Omic Era*, DOI 10.1007/978-1-4614-8370-0_12,

Keywords Bioinformatics • Sequence analysis • Functional annotation • Comparative mapping • Evolutionary biology • Food legume • Molecular markers • Sequence database • In silico analysis

Introduction

The complete genome has been sequenced in three legume species namely, *Medicago truncatula*, *Lotus japonicus* and soybean (*Glycine max*) (Bertioli et al. 2009; Cannon et al. 2009; Sato et al. 2008; Zhu et al. 2005; Schmutz et al. 2010). Among these, *M. truncatula* is considered as model species, and is taxonomically more related to cool-season legumes such as pea, lentil, faba bean, and chickpea (Bordat et al. 2011). Integrating the genomic and biological knowledge from model legumes to other economically important cool-season pulse crops, e.g., pea, lentil, and chickpea, warm-season food legumes, e.g., peanut and common bean, and forage legumes, e.g. alfalfa and clover, will provide a major opportunity for advancing their genomic resources (Young et al. 2005; Young and Udvardi 2009; Varshney and May 2012). For example it can foster gene identification in such species, which are less noticeable due to their large genomes (Gepts et al. 2005). Sequencing of other legumes, including common bean (Ramírez et al. 2005; David et al. 2008) is progressing rapidly and draft genome sequences of some of them like pigeonpea (Varshney et al. 2009, 2011; Singh et al. 2012) and chickpea (Garg et al. 2011; Varshney et al. 2013) are already available.

Various genome sequencing projects have produced a wealth of sequence data, which need to be properly analysed to enable prediction of the potential functional elements, genes and transcription factors. Rapid progress has been made to develop bioinformatics tools and databases for such analyses as well as for understanding of the various features of the sequenced genome (Kushwaha et al. 2008; Dutt et al. 2010; Kumari et al. 2010). Similarly, *in-silico* comparative genomics provides a great opportunity in unravelling the behaviour of genes and genomes (Udvardi 2002; Kushwaha et al. 2012). Comparative genomics uses information about signature parts at the gene level and syntenic relation at the genome level to understand the structure and function of a newly sequenced genomes, as well as to deduce its evolutionary relationships (Goffard and Weiller 2006). Gene hunting is another important application of comparative genomics to investigate coding and non-coding functional elements of the genome (Yadav et al. 2007; Kushwaha et al. 2011). It attempts to discover both similarities and differences in the genes, proteins, RNA, and regulatory regions of different organisms to infer structural and functional relationships. Comparative genomics is now focusing on discovery of regulatory regions and siRNA molecules in the genome. The available biological datasets in web repository databases allow for comparative analysis and real data validation with the existing datasets. Different databases maintained by a data model like NCBI are integrated with each other to enable their effective utilization. The experimental datasets thus give us opportunities to understand the functional and biological roles

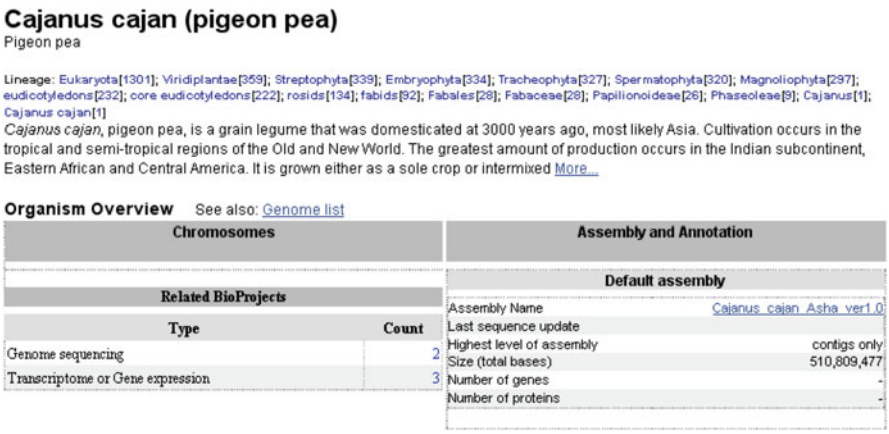
Table 12.1 Important biological databases related to legumes

Database name	URL
NCBI	http://www.ncbi.nlm.nih.gov/
DNA Data Bank of Japan DDBJ	www.ddbj.nig.ac.jp/
EMBL	www.ebi.ac.uk/em
United States Dry Bean Council (USDBC)	http://www.usdrybeans.com/
International Legume Database and Information Service (ILDIS)	http://www.ildis.org/
Legumes information System	http://www.comparative-legumes.org/
Legume "Phylo-informatics" dbase	http://www.public.asu.edu
Food Legume genome database	http://www.gabcsfl.org/
SoyBase	http://soybase.org
Medicago truncatula	http://www.medicago.org/
Illustrated Legume Genetic Resources Database	www.gene.affrc.go.jp
SSR Database of legumes	http://intranet.icrisat.org/gt1/ssr/ssrdatabase.html
Bioinformatics resources for legume researchers	http://www.legumes.org/
Chinese Legume Database and Information Service (CLDIS)	http://cldis.ibcas.ac.cn/
LegumeTFDB	http://legumetfdb.psc.riken.jp/
Lotus japonicus	http://www.kazusa.or.jp/lotus/
Phytozome v7.0	http://www.phytozome.net/
Chickpea Transcriptome Database	http://59.163.192.90:8080/ctdb/
Chickpea Root EST Database	http://www.icrisat.org/what-we-do/biotechnology/Cpest/home.asp
Gramene	http://www.gramene.org/
GmGDB	http://www.plantgdb.org/GmGDB/
Lotus japonicus genome DB	http://www.kazusa.or.jp/lotus/
Legume Information System	http://www.comparative-legumes.org/
Common Bean Database	http://jeff.ifxworks.com/Legume/common_bean.html

of unknown genes/proteins from different legumes. The availability of different biological databases related to legumes provides valuable information resource for research and analysis (Table 12.1). However, the main aim of bioinformatics is the identification of regulatory mechanisms and function of genomes and their evolution (Marla and Singh 2012).

Bioinformatics for Legume Genome Annotation

Sequencing determines the primary structure of an unbranched biopolymer. The elements with the associated function can be predicted by using DNA/protein sequences. Sequencing of a genome is a complicated and typical task that uses DNA sequencing to determine the order of nucleotides in small DNA fragments that together make up the genome. The first generation DNA sequencing was performed



Cajanus cajan (pigeon pea)
Pigeon pea

Lineage: Eukaryota[1301]; Viridiplantae[359]; Streptophyta[339]; Embryophyta[334]; Tracheophyta[327]; Spermatophyta[320]; Magnoliophyta[297]; eudicotyledons[232]; core eudicotyledons[222]; rosids[134]; fabids[92]; Fabales[28]; Fabaceae[28]; Papilionoideae[26]; Phaseoleae[9]; Cajanus[1]; Cajanus cajan[1]

Cajanus cajan, pigeon pea, is a grain legume that was domesticated at 3000 years ago, most likely Asia. Cultivation occurs in the tropical and semi-tropical regions of the Old and New World. The greatest amount of production occurs in the Indian subcontinent, Eastern African and Central America. It is grown either as a sole crop or intermixed More...

Organism Overview See also: Genome list

Chromosomes	Assembly and Annotation

Related BioProjects

Type	Count
Genome sequencing	2
Transcriptome or Gene expression	3

Default assembly

Assembly Name	Cajanus_cajan_Asha_ver1.0
Last sequence update	
Highest level of assembly	contigs only
Size (total bases)	510,809,477
Number of genes	-
Number of proteins	-

Fig. 12.1 An Example of pigeonpea (*C. cajan*) genome sequence deposited in NCBI by a group of Indian scientists [Reprinted from Singh N. K., Gupta D. K., Jayaswal P. K., Mahato A.K., Dutta S., Singh S., Bhutani S., et al. (2012) The first draft of the pigeonpea genome sequence. J. Plant Biochem Biotechnol 21: 98–112 with permission from Springer Science+Business Media]

by using the chain termination method developed by Frederick Sanger and co-workers (Sanger and Coulson 1975; Sanger et al. 1977). This technique uses sequence-specific termination of a DNA synthesis reaction using modified nucleotide substrates. However, new sequencing technologies such as pyrosequencing are gaining an increasing share of the sequencing work and the next generation DNA sequencers that achieve sequencing by synthesis are based on this approach. These sequencer do not require *in vivo* library construction, are faster and much cheaper to use; they are being used for rapid genome sequencing. An example of nearly completed *C. cajan* genome sequenced by a group of Indian scientists using the second generation DNA sequencers is depicted in Fig. 12.1.

After completion of the full genome sequence, it is necessary to assemble and annotate new sequences. In fact, genome assembly is a very difficult computational task owing to large numbers of identical sequences (repeats) found in genomes. These repeats can be of thousands of nucleotides in length, and some of them may occur in a number of different locations. In a shotgun sequencing project, the entire DNA from a source (usually a single organism, ranging from a bacterium to a mammal) is first fragmented into millions of small pieces. These pieces are then "read" by automated sequencers, and each read can be up to 1,000 nucleotides long. A genome assembly algorithm works by taking all the reads and aligning them with one another, to detect all the places where two of the reads are overlapping. These overlapping reads can be merged together to form a contig and then linking information of contigs is used to create scaffolds. Subsequent to this, scaffolds are positioned along the physical map of the chromosomes.

Most of the assembler tools and packages were developed by different research groups, e.g., short oligonucleotide analysis package and *de novo* assembly tools were developed by Beijing Genomics Institute (BGI).

Table 12.2 Bioinformatics softwares available for genome annotation and *de novo* assembly

Application	Available tools
Genome annotation	TRF, Repeat Masker, Genescan, BGF, InterproScan etc.
De-novo assembly	SOAP *de-novo*, AbySS, Velvet etc.
Genome resequencing analysis	SOAPsnp\SOAPsv\SOAPInDel, SAMtools, BreakDancer, VarScan etc.

In genome annotation one can elucidate the biological information based on assembled genome sequences. In this process, called "gene prediction", one can identify functional elements in the genome and generate biological information about these elements. The genome annotation is done by the methods prescribed by Kawaji and Hayashizaki (2008). The basic level of genome annotation can be done using Basic Local Alignment Search Tool BLAST to find out similarities and differences. However, nowadays more and more additional information is added to the annotation platform. The complete annotated genome data are deposited in different biological databases, i.e., NCBI, DDBI, Phytozome, Ensembl and EMBL. These databases use genome context information, experimental datasets, and integrations of tools and resources to provide gene and genome annotations through their subsystems approach. Sequence Assembly AMOS tool can be used for manipulation with sequence files. AMOS tool is currently maintained by University of Maryland. CABOG is a tool that assembles large genomic DNA sequences produced by whole-genome shotgun sequencing. Some important annotation tools like Apollo, BLAST, Parser, MATLAB, Bioconductor package in R, Artemis and AAT tool are available. Manatee is a web-based gene evaluation and genome annotation tool for visualization, modification and storage for genomes. PASA can be used as eukaryotic genome annotation tool that exploits spliced alignments of expressed transcript sequences to gene model. Several bioinformatics tools are available for annotation, genome sequence alignment, *de novo* assembly, sequence alignments, evolution and RNA sequence analysis; some of these tools are listed in Table 12.2.

Hiremath et al. (2011) carried out a large-scale transcriptome analysis in chickpea (*C. arietinum* L.) using next generation sequencing technologies such as, Roche 454 and Illumina/Solexa. They determined a total of 103,215 tentative unique sequences (TUSs) and assigned functions for 49,437 (47.8 %) of the TUSs. Comparison of the chickpea TUSs with the *M. truncatula* genome assembly (Mt 3.5.1 build) resulted in 42,141 aligned TUSs with putative gene structures (including 39,281 predicted intron/splice junctions). These TUSs were also used to identify 728 SSR, 495 SNP, 387 conserved orthologous sequence (COS) markers, and 2,088 intron-spanning region (ISR) markers. Similarly, transcriptome assembly has been done in pigeonpea by Kudapa et al. (2012) referred to as CcTA v2, comprised 21,434 transcript assembly contigs (TACs) and 77.5 % TACs (16,622 TACs) of the total could be mapped on to the soybean genome. Based on knowledge of intron junctions, so far 10,009 primer pairs were designed from 5,033 TACs for amplifying intron spanning regions (ISRs). By using *in silico* mapping of BAC-end-derived

SSR loci of pigeonpea on the soybean genome as a reference, putative mapping positions at the chromosome level were predicted for 6,284 ISR markers, covering all the 11 pigeonpea linkage groups. The transcript assembly and markers developed will provide a useful resource for basic and applied research for genome analysis and crop improvement in chickpea and pigeonpea.

ORFs and their localization, gene structure optimization, coding region identification and location of regulatory motifs explain the complete organization of gene family with their associated functions. Identification of gene family is a better approach to investigate the various types of members related to each other and the manner in which they have evolved (Thornton and DeSalle 2000). Availability of EST datasets for a genome gives a better understanding of transcripts with tissue-specific expression. Based on bioinformatics tools and databases any one can compare biological experiment datasets with any query sequence. *In-silico* based approaches utilize information from expressed sequence tags and proteins, often derived from mass spectrometry, to improve genomic annotations. A variety of software tools have been developed to help scientists in their quest for gene and genome annotations. Identification of gene locations and the sites of other genetic control elements are often described as the biological "parts list" for the assembly of an organism. Scientists are still at an early stage of delineating this "parts list" and in understanding how all the parts fit together and work together. Gene and genetic control elements investigation can be done using publicly available biological databases and tools accessible *via* the web and other electronic means. Some statistical tools are available for the analysis of deep sequencing like ANDES Tools and DAG chainer that computes chains of syntenic genes within complete genome sequences. DNA sequence analysis tools include k-mer tool, ESTmapper, Snapper mapping reads and ATAC are available for aligning genomes. For rapid aligning of the entire genomes, a software MUMmer, can be used.

Bioinformatics for Sequence Analysis

In bioinformatics, sequence analysis refers to the process of subjecting a DNA, RNA or protein sequence using analytical methods and algorithms to understand its features, function, structure, or evolution. Methodologies used are biological database mining, comparative analysis and sequence alignment. With the development of statistical algorithm, matrices based tools for prediction of gene and protein sequences, the rate of addition of new sequences to the databases has increased exponentially. Such a collection of sequences does not, by itself, increase the scientist's understanding of the biology of organisms. However, comparing these new sequences to those with known functions is a key way of understanding the biology of an organism from which the new sequence comes. Thus, sequence analysis can be used to assign functions to genes and proteins by a study of the similarities between the compared sequences. Nowadays, there are many tools and techniques are available that provide the sequence comparisons (sequence alignment) and analyze the alignment

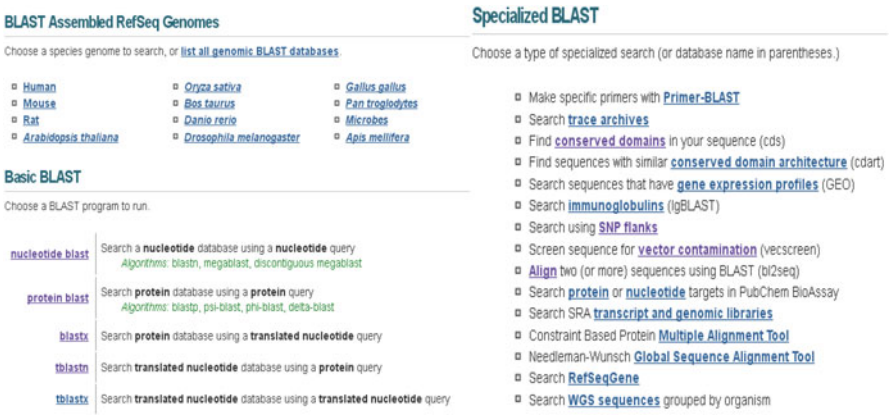


Fig. 12.2 A page showing basic local alignment search tool (BLAST; http://blast.ncbi.nlm.nih.gov/)

of a product to understand its biology. Sequence analysis in molecular biology includes a wide range of applications, some of which are listed below.

1. Comparison of different sequences in order to detect similarities among them and, often, to infer if the sequences are related (homologous).
2. Identification of intrinsic features of the different sequences, such as active sites, post-translational modification sites, gene structures, reading frames, distributions of introns and exons and the regulatory elements.
3. Identification of sequence differences and variations such as point mutations and single nucleotide polymorphisms (SNPs) in order to develop the genetic markers.
4. Unraveling the evolutionary process and assessment of genetic diversity of the sequences and the organisms.
5. Identification of molecular structure from sequence data alone.

Sequence analysis is based on sequence alignment, i.e., comparison between query and subject sequences, in which two or more sequence sets can participate. Alignment between two sequences is called pairwise alignment, and alignment between more than two sequences is called multiple sequence alignment. Two methods are used for searching for a series of identical or similar characters in the sequences to find out similarities and dissimilarities within sets of sequences; these are called global and local alignments. Global alignment finds the best alignment across the whole length of two sequences and forces alignment in such regions that show differences. Local alignment finds regions of high similarity in parts of the participating sequences, and concentrates on regions of high similarity. Basic local alignment search tool (BLAST) is an example of local alignment (Fig. 12.2). Mainly five flavors of Basic BLAST are available for comparison of the query with the subject for sequence. In case of protein query sequence, one can use BLASTp and tBLASTn. In case of nucleotide query sequence, any one of the BLASTn, BLASTx and tBLASTx can be used. Other specialized blasts are also available for conserved domain detection, SNP detection, global sequence alignment, etc.

Gene Identification and Characterization Using Comparative Genomics/Proteomics

In computational biology gene hunting or gene prediction refers to the process of identifying the regions of genomic DNA that function as genes, i.e., encode proteins or various types of RNA molecules, or as other functional elements like regulatory regions. Gene finding is one of the first and most important steps in understanding the genome of a species once it has been sequenced. Earlier "gene finding" was based on cumbersome experiments on living cells and organisms. But the availability of comprehensive genome sequences and powerful computational resources have greatly facilitated gene finding, and some of the tools and database servers dedicated to gene prediction are listed in Table 12.3.

Genome sequence of "Asha" variety of pigeonpea was obtained using GS-FLX Phase D chemistry and the GS-FLX Titanium chemistry and reads were assembled

Table 12.3 A list of some important gene prediction servers

Name	Description/function
ATGpr	Identifies translational initiation sites in cDNA sequences
AUGUSTUS	Predicts genes in eukaryotic genomic sequences
BGF	Hidden Markov model based *ab initio* gene prediction program
EUGENE	Gene hunting for *Arabidopsis thaliana*
FRAMED	Finds genes and frameshift in G + C rich prokaryotic sequences
GENIUS	For linking predicted genes in complete genomes to known protein 3D structures
GENEID	Signal, exon and gene prediction server
GENEPARSER	Detect intron and exon regions in DNA sequence
GeneMark	Family of gene prediction programs
GeneMark.hmm	A gene prediction program for prokaryotes and eukaryotes
GeneTack	Prediction of genes with frameshifts in prokaryotic genomes
NIX	Web tool gene prediction based on combining results from different programs
GLIMMER	For finding genes in microbial DNA
VEIL	Hidden Markov model for finding genes in vertebrate DNA Server
Splice Predictor	Identifies potential splice sites in (plant) pre-mRNA using Bayesian methods
GENESCAN	For finding genes using Fourier transform
FGENESH	The fastest and most accurate *ab initio* gene prediction program
NNPP	Promoter prediction by neural network
NNSPLICE	Splice site prediction using neural network method
GENOMESCAN	Predicts locations and exon-intron boundary in genomic sequences
ORF FINDER	A graphical analysis tool for open reading frame prediction
GrailEXP	Predicts exons, genes, promoters, poly-As, CpG islands and repetitive elements within DNA sequences
EuGène	Gene finder for eukaryotic system exploits probabilistic models for discriminating coding from non-coding sequences to discriminate effective splice sites from false splice sites

using "Newbler GS De Novo assembler version 2.5.3" that compares all sequence reads pairwise and reads with overlaps are joined into contigs (Singh et al. 2011). An average of all aligned reads at a specific nucleotide position is used to determine the consensus sequences for a contig, and overlapping contigs are finally merged to make scaffolds. The finished sequence was passed through fgenesh tool of Molquest software using *Arabidopsis thaliana* gene models as a reference. Predicted genes with size of >500 bp were BLAST-searched against the NCBI database, and the search output was processed using BLAST Parser software and gene annotations were manually curated and categorized based on function. Singh et al. (2012) were able to predict a total of 59,515 genes with the largest size of 11,523 bp and the smallest gene size of 501 bp of these 47,004 were protein coding genes of which 1,213 were related with plant defense and 152 were involved in abiotic stress tolerance.

Comparative phylogenetic studies within the legume family revealed high syntenic relationships between sequenced legumes and other important legumes (Wojciechowski et al. 2004), e.g. between *Medicago truncatula* and pea (Kaló et al. 2004), and common bean and soybean (Lee et al. 2001), but limited synteny is also reported to be present among other legumes, e.g., between cool-season and warm-season legumes (Zhu et al. 2005). Whole genome sequencing of some important legumes is likely to be completed in the near future, and this will facilitate a comprehensive assessment of synteny. Comparative genomics for synteny studies can accelerate exploitation of genomic resources, and facilitate more rapid progress in research efforts in an efficient and cost-effective manner. A detailed study of the syntenic relationships is a critical issue to be addressed for better allocation of genomic information from sequences of model legumes to other legumes and to other crop species. Based on conservation of synteny between pigeonpea and soybean genomes, Singh et al. (2012) found that chromosomes 1, 3, 4 and 9 of pigeonpea showed the maximum conservation with chromosomes 2, 5, 7, 8, 12, 13, 15 and 17 of soybean. Chromosome 1 of pigeonpea showed the highest number of matches with chromosomes 8 and 5 of soybean. Similarly, chromosome 2 of pigeonpea showed the maximum number of hits with chromosomes 19 and 10 of soybean. Pigeonpea chromosome 3 showed the maximum number of hits with chromosomes 13 and 15 of soybean, pigeonpea chromosome 4 showed the maximum number of hits with chromosomes 12 and 13 of soybean, chromosome 5 showed the highest number of matches with chromosomes 13, 12 and 17 of soybean, chromosome 6 showed the maximum number of matches with chromosomes 9 and 3 of soybean, chromosome 9 showed maximum number of matches with chromosomes 2, 12, 3, 11 and 16 of soybean, chromosome 10 showed the maximum number of hits with chromosomes 18, 17 and 2 of soybean, chromosome 11 showed the maximum numbers of hits with chromosomes 14 and 18 of soybean, and chromosome 7 showed maximum number of hits with chromosomes 10 and 20 of soybean, while chromosome 8 of pigeonpea showed minor synteny with chromosomes 13 and 14 of soybean. However, Singh et al. (2012) concluded that the overall synteny between the genomes of pigeonpea and soybean was only to a limited extent.

Bioinformatics for Computational Evolutionary Biology

The phylogenetic tree (phylogeny) is textual and visual representation that describes evolutionary relationships among various groups of organisms or among a family of related nucleotide or protein sequences and other entities based upon similarities and differences in their physical and genetic characteristics. In such a study, one can use morphological features (e.g., shape, size, length, etc.) and molecular data (e.g., DNA and protein sequences). The taxa/entities joined together in the tree are implied to have descended from a common ancestor. Phylogenetic trees are useful in fields of bioinformatics, systematics and comparative biology. There are rooted and unrooted types of tree inferences and main approaches for phylogeny reconstruction, i.e., distance based methods, topology search methods and Bayesian methods. Some phylogenetic tree terminologies are shown in Fig. 12.3.

A rooted phylogenetic tree defines common ancestor of all the entities at the leaves of the tree, i.e., the operational taxonomic units (OTUs). One example showing root based phylogenetic classification of Toll interleukin 1 receptor (TIR) domain among different organisms depicts the way this family might have been derived during evolution (Fig. 12.3). Phylogenetic relationships among genes can help to predict the genes that might have similar function e.g. *ortholog detection*.

TIR domain is mainly involved in plant immune responses against various pathogens. An example of Toll/interleukin-1 receptor classification is provided here TIR domain for *C. cajan* was used for find out similar homologues in different organisms using basic local alignment search tool (BLAST). Selected homologues from different species were used for multiple sequence alignment and phylogenetic

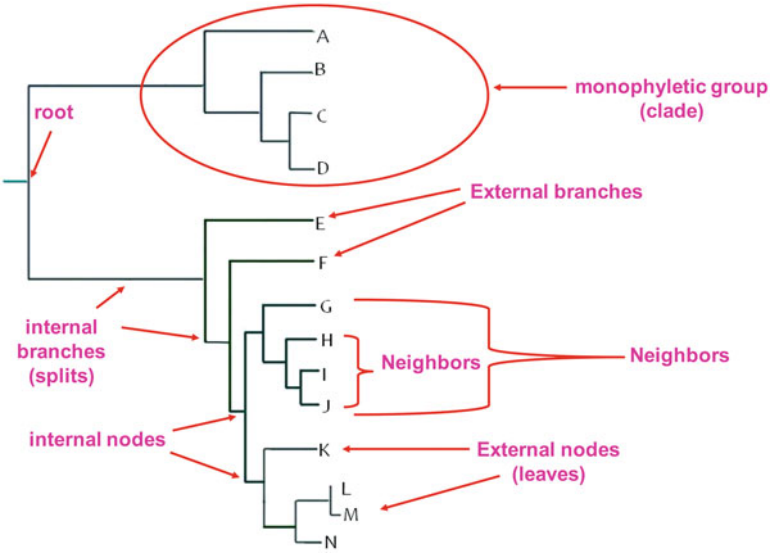


Fig. 12.3 Figure showing phylogenetic tree terminologies

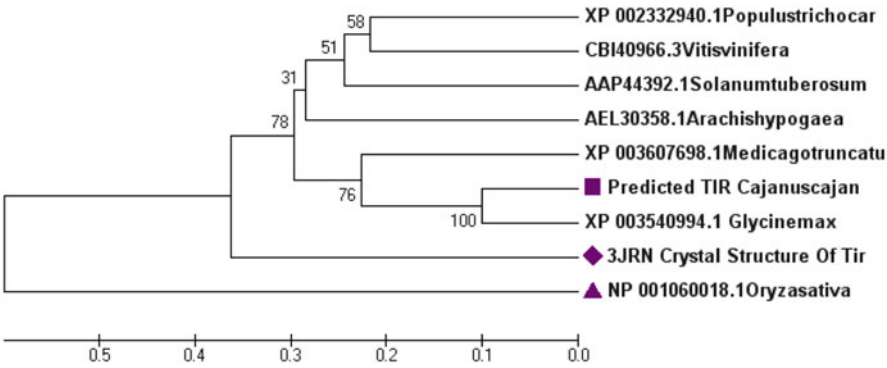


Fig. 12.4 Example of rooted tree of TIR domain homologues from *C. cajan* with six other plant species (Singh et al., unpublished data)

classification. ClustalW tool was used for multiple sequence alignment and for tree classification, MEGA tool was used to find out the best tree topology. Figure 12.4 shows the rooted inferences of selected sequences of TIR domains from seven different plant species (*Populus*, *Vitis*, *Solanum*, *Arachis*, *Medicago*, *Glycine*, *Cajanus* and *Oryza*). Interestingly, it was found that TIR, *Oryza* spp. forms an outer group, while the remaining six TIR domains are much more closely related this may be expected because *Oryza* is a monocot.

The identified TIR domain from *C. cajan* was further used to determine the number of TIR loci present in the *Cajanus* genome, and a total of 148 TIR domains have been successfully identified based on the available datasets of *C. cajan* genome sequence (Taxid: 3821). Figure 12.5 shows an unrooted tree depicting the various TIR domains derived from Cajanus genome itself. Unrooted trees specify relationships but they do not depict the evolutionary path. For phylogenetic study, different online and offline softwares are available (Table 12.4). Legume diversity and evolution in a phylogenetic context has been reviewed earlier by Doyle and Luckow (2003).

In-Silico Analysis for Gene Expression Data

An expressed sequence tag (EST) is a short, ordinarily, terminal sequence of a cDNA sequence. Thus an EST results from one-shot sequencing of a cloned mRNA, i.e., several hundred base pairs of sequence starting from an end of a cDNA sequence. The cDNAs used for EST generation are typically individual clones from a cDNA library. ESTs may be used to identify gene transcripts; they are instrumental in gene discovery and gene sequence determination. The identification of ESTs has proceeded rapidly, and ~73 million ESTs are now available in the public database GenBank. The dbEST is a division of Genbank established in 1992, and the data in dbEST is directly submitted by laboratories worldwide. Based on EST

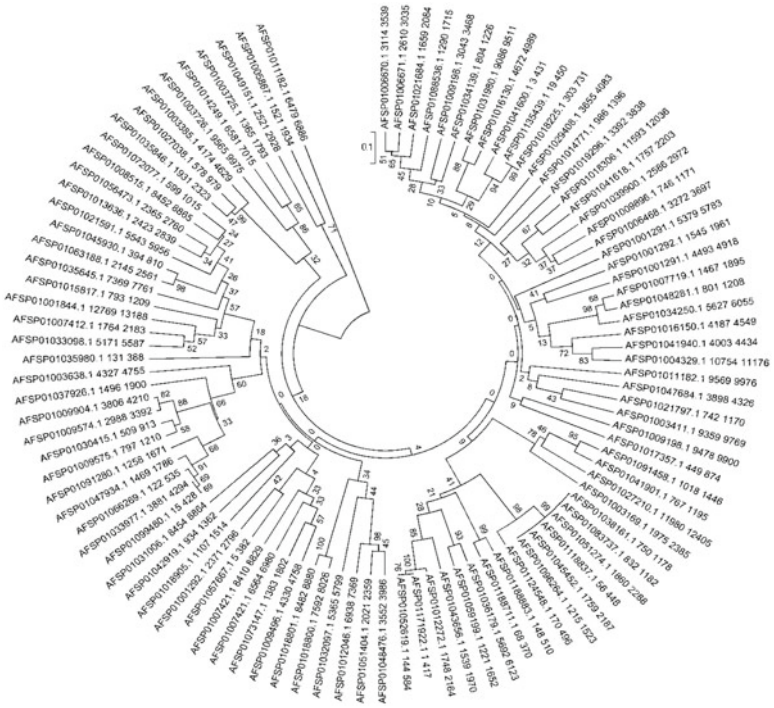


Fig. 12.5 Example of unrooted tree of identified TIR domains from *C. cajan*

Table 12.4 Tools and servers for multiple sequence alignment and phylogenetic analysis

Tools and server	URL
ClustalW2	http://www.ebi.ac.uk/Tools/msa/clustalw2/
CLUSTALW	http://www.genome.jp/tools/clustalw/
MEGA	http://www.megasoftware.net/
T-Coffee	http://www.ebi.ac.uk/Tools/msa/tcoffee/
PHYLIP	http://evolution.genetics.washington.edu/phylip.html
The PhylOgenetic Web Repeater (POWER)	http://power.nhri.org.tw/power/home.htm
BlastO	http://oxytricha.princeton.edu/BlastO/
BIONJ	http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::bionj
DendroUPGMA	http://genomes.urv.cat/UPGMA/
PhyML	http://www.atgc-montpellier.fr/phyml/binaries.php
Evolutionary Trace Server (TraceSuite II)	http://mordred.bioc.cam.ac.uk/~jiye/evoltrace/evoltrace.html
Phylogeny.fr	http://www.phylogeny.fr/
Mesquite	http://mesquiteproject.org/mesquite/mesquite.html
Winboot	http://archive.irri.org/science/software/winboot.asp

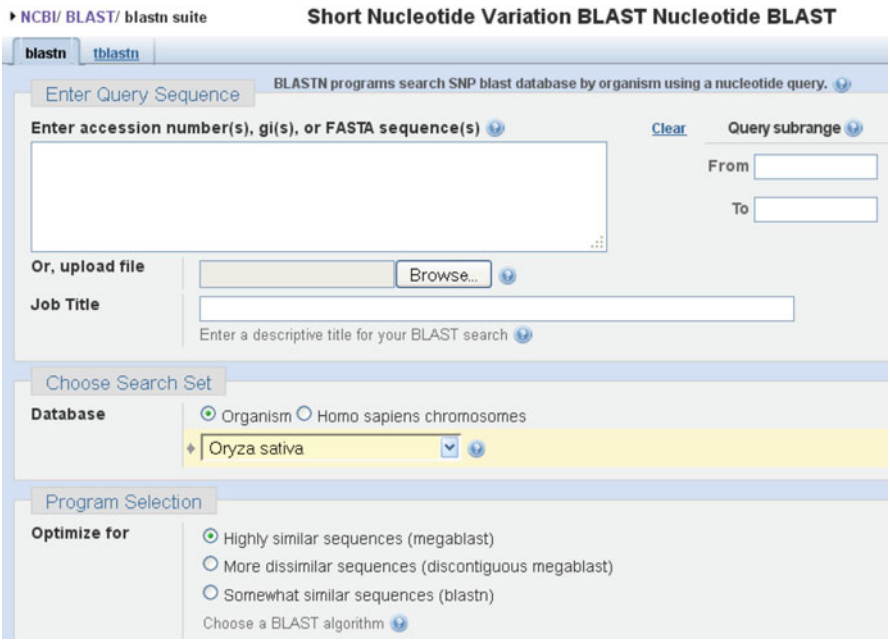


Fig. 12.6 Short nucleotide variation BLAST page

datasets any one can determine the gene function based on expression datasets. ESTs contain enough information to permit the design of precise probes for DNA microarrays that can be used to determine the gene expression. For expression microarray data analysis normalization and management, one can use Ginkgo (Comparative Genomic Hybridization package). TM4 and Magnolia packages are also designed for microarray data management for researchers who use PFGRC microarrays. The programme SNP Filter Scripts can be used to identify and detect false positive SNP calls that are present in raw data from affymetrix gene chip resequencing arrays. There are several other tools freely available, including MAGIC, CLUSFAVOUR, etc. for microarray data analysis. Short nucleotide variation analysis server is also available for this type of study (Fig. 12.6).

Bioinformatics in Legume Nutritional Genomics

By manipulating the promoter region of seed-specific protein encoding genes one can improve the nutritional quality of any crop species. Bioinformatics tools can play a major role in the study of the promoter region of genes and for identification of *cis*-acting elements or *cis-regulatory* elements. A *cis*-acting element is a

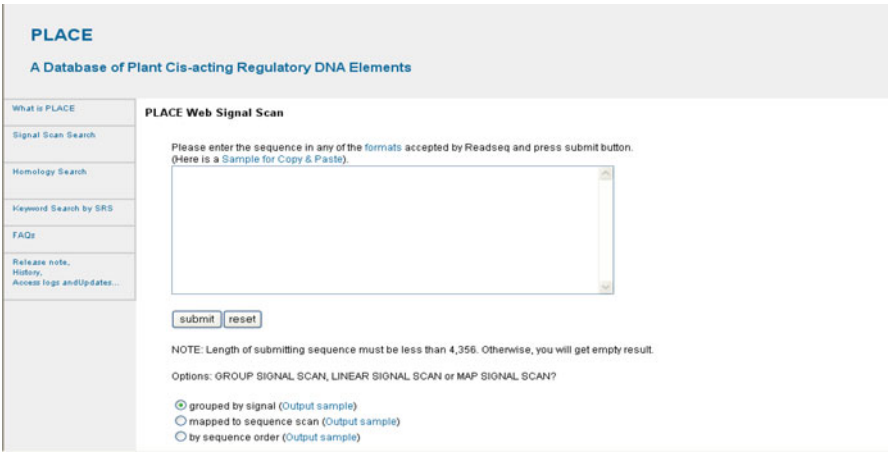


Fig. 12.7 Plant *cis*-acting elements prediction server (PLACE; http://www.dna.affrc.go.jp/PLACE/)

region of DNA or RNA that regulates the expression of genes located in the same chromosome. This term is derived from the Latin word *cis*, which means "on the same side as". The *cis*-regulatory elements are often binding sites for one or more *trans-acting* factors. These *cis*-elements may be located upstream of the coding sequences of the concerned genes, i.e., in the promoter region or even further upstream, in an intron, or downstream of the gene's coding sequence. In molecular biology and genetics, a transcription factor (sometimes called a sequence-specific DNA-binding factor) is a protein that binds to specific DNA sequences, thereby controlling the flow of genetic information (or transcription) from DNA to mRNA. Transcription factors perform this function alone or with other proteins in a complex, by promoting (as an activator)/or blocking (as a repressor) the recruitment of RNA polymerase to transcribe specific genes. Therefore, identification of potential *cis*-acting elements can help in improving the nutritional quality of seeds of plant species, and/or other traits of economic/agronomic value.

Databases of plant *cis*-acting regulatory elements like PlantCare and PLACE can be used as a portal for *in-silico* analysis of promoter sequences of plant genes (Fig. 12.7). Yadav et al. (2007) successfully identified the seed storage protein promoter specific *cis*-acting elements in cloned and sequenced promoter regions of seed storage protein genes from different cultivars of wheat, rice and oat. A database containing collection of proximal promoter sequences for RNA polymerase II with experimentally determined transcription start-sites from various plant species is available on server PlantProm DB. For retrieval and investigation of transcription factor associated genes PlnTFDB (plntfdb.bio.uni-potsdam.de/) and PlantTFDB (http://planttfdb.cbi.pku.edu.cn/) are important databases. In addition, species transcription factor databases are also available online (Fig. 12.8).

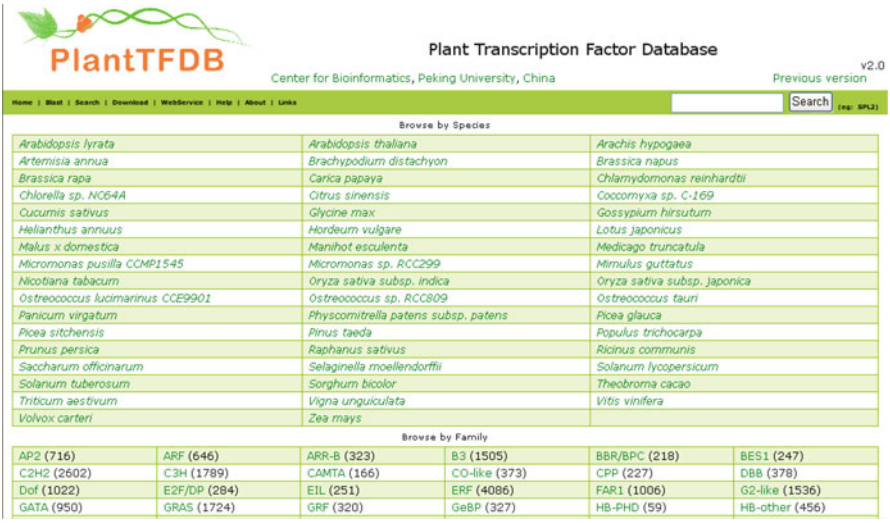


Fig. 12.8 Plant transcription factor database PlantTFDB (http://planttfdb.cbi.edu.cn/)

Prediction for Function of Protein Sequences

In the prediction of the function of a protein sequence of interest, structural visualization, 3D prediction, classification and structural alignment play important roles. In this connection homology modeling, threading and *ab-initio* prediction methods can be used for protein structure prediction. Homology modeling (comparative modeling) is a process for constructing an atomic-resolution model of the "target" protein using an experimental three-dimensional structure of a related homologous protein (the "template") derived by NMR, X-ray techniques. Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence. It has been shown that protein structures are more conserved than the amino acid sequences amongst homologues, but sequences falling below 20 % sequence identity can have very different structures. For homology modeling, threading and *ab-initio* prediction several servers are available in public domain (Table 12.5). Some commercial software like MOE, Schrödinger and Discovery Studio can also be used for protein modeling and simulation. For *Ab-initio* or *de-novo* protein modeling one can use I-TASSER and ROBETTA, which are freely available. Based on different protein modeling servers, one can predict the three dimensional structure of the target protein.

Table 12.5 List of servers for homology modeling, threading and *ab-initio* based structure prediction

Server name	Description	URL
SWISS-MODEL	Automated protein structure homology-modeling server	http://swissmodel.expasy.org/
ModBase	Comparative modeling based on three-dimensional protein models. The models are derived by ModPipe, an automated modeling using PSI-BLAST and MODELLER	http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi
I-TASSER	Model is built based on multiple-threading alignments by LOMETS and iterative TASSER simulations	http://zhanglab.ccmb.med.umich.edu/I-TASSER/
LOMETS	3D model prediction by collecting high-scoring target-to-template alignments using threading programs (FUGUE, HHsearch, MUSTER, PPA, PROSPECT2, SAM-T02, SPARKS, SP3)	http://zhanglab.ccmb.med.umich.edu/LOMETS/
ESyPred3D	Homology modeling web by combining, weighting and screening the results of several multiple alignment programs using the modeling package MODELLER	http://www.fundp.ac.be/sciences/biologie/urbm/bioinfo/esypred/
3D-Jigsaw	Automated system to build three-dimensional models for proteins based on homologues of known structure	http://bmm.cancerresearchuk.org/~3djigsaw/
HMMSTR/Rosetta	Predicts the structure of proteins from the sequence: secondary, local, super secondary, and tertiary. Provided by the Depts. of Biology and Computer Science, Rensselaer Polytechnic Institute	http://www.bioinfo.rpi.edu/bystrc/hmmstr/server.php
Geno3D	Protein three-dimensional structure using comparative protein structure modeling by spatial restraints (distances and dihedral) satisfaction	http://geno3d-pbil.ibcp.fr/cgi-bin/geno3d_automat.pl?page=/GENO3D/geno3d_home.html
VADAR (Volume, Area, Dihedral Angle Reporter)	Quantitatively and qualitatively assess protein structures determined by 3D-threading or homology modelling	http://vadar.wishartlab.com/
ResProx (Resolution-by-proxy or Res(p))	A web server that predicts the atomic resolution of NMR protein structures using only PDB coordinate data as input	http://www.resprox.ca/
Robetta	*Ab initio* fragment assembly	http://robetta.bakerlab.org/

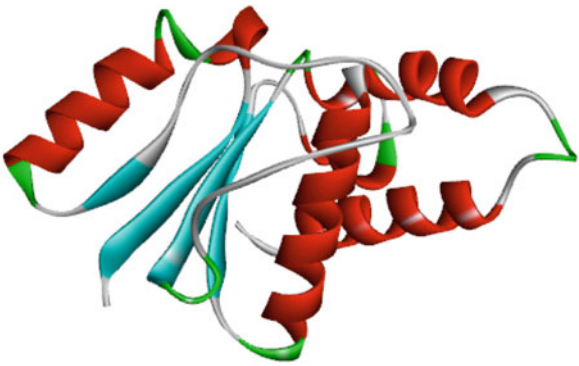


Fig. 12.9 Structure of TIR domain (PM0078097) from *C. cajan* developed using TIR domain structure from *Arabidopsis thaliana* (3JRN) based on homology modelling [Courtesy of Vinay Kumar Singh]

Qualitative and Quantitative Study of Predicted Models

Finally, predicted 3D models can be subjected to a series of tests for assessing their internal consistency and reliability. The Quality of the model can be checked with verify3D [http://nihserver.mbi.ucla.edu/Verify_3D/], Errat [http://nihserver.mbi.ucla.edu/ERRATv2/] etc. The stereochemical properties based on backbone conformation can be evaluated by inspection of Psi/Phi/Chi/Omega angle using Ramachandran plot of PDBSum database [http://www.ebi.ac.uk/pdbsum/], RAMPAGE [http://mordred.bioc.cam.ac.uk/~rapper/rampage.php] etc. Quantitative analysis can be done using accessible surface area prediction using Volume Area Dihedral Angle Reporter [VADAR; http://vadar.wishartlab.com/]. Standard bond lengths and bond angles of the model can be determined using WHAT IF [http://swift.cmbi.ru.nl/whatif/]. ResProx (Resolution-by-proxy; http://www.resprox.ca/) can be used for quality and quantity measurements at resolution level. For example, we have successfully predicted 3D model of toll-like interleukin receptor (TIR) domain of *R* genes from *C. cajan* using comparative homology modeling and the best evaluated model has been deposited to Protein Model DataBase (PMDB; http://mi.caspur.it/PMDB/) (Fig. 12.9).

Integrated Bioinformatics Tools

Some integrated tools like MEME and MAST are useful servers for motif elucidation (Fig. 12.10). For protein functional elucidation and characterization, one can use INTERPROSCAN, PROSITE, PFAM and PRODOM etc. (Fig. 12.11). SWISSPROT, DBSNP and SNP flanks tools and databases can be used for SNP/variant detection. An example of signature part of toll-like interleukin receptor domain from *C. cajan* is given in Fig. 12.12.

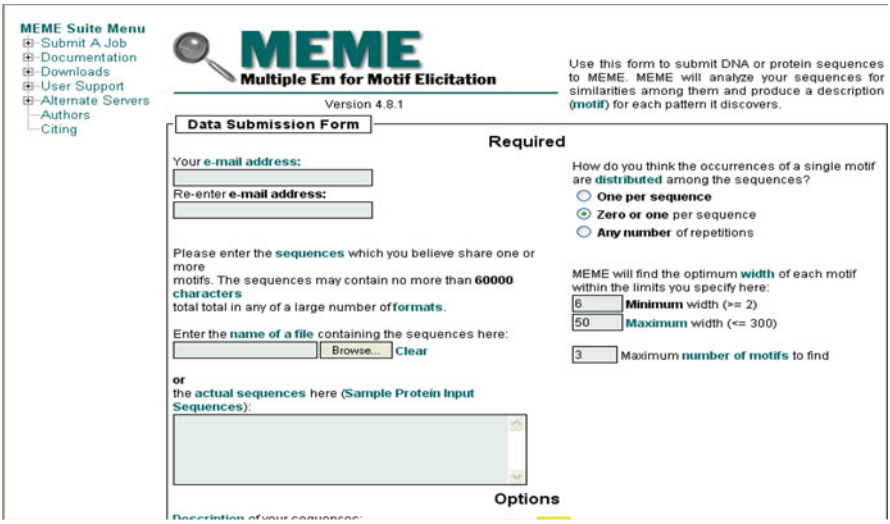


Fig. 12.10 A server to discover motifs (highly conserved regions) in groups of related DNA or protein sequences

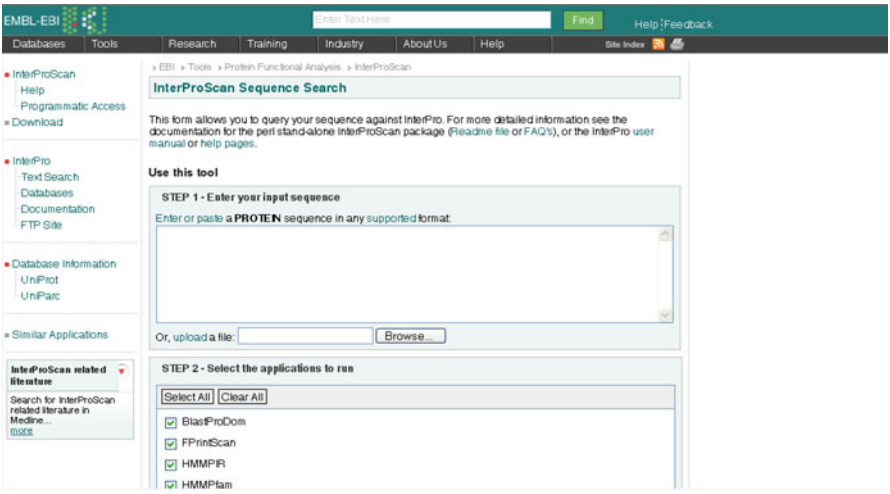


Fig. 12.11 Server for protein functional elucidation based on domain and signature motifs

Molecular Docking

In bioinformatics, molecular docking is a method that predicts the possible orientation of one molecule in relation to a second when the two are bound to each other to form a stable complex. The knowledge of the possible orientations in turn, can be

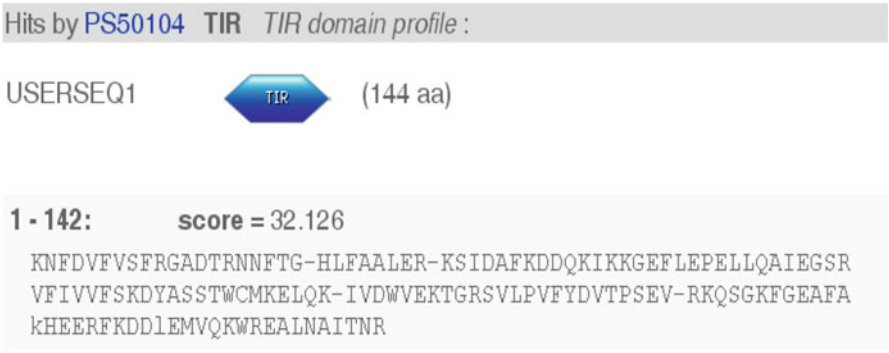


Fig. 12.12 Toll-like interleukin receptor domain form *C. cajan*

used to predict the binding affinity between the two molecules using energy scoring functions. Using molecular docking approach, one can predict the binding orientation with energy total and energy shape of a ligand (small molecule) to its protein target (receptor) to predict the affinity and activity of the small molecule. The interaction between ligand and receptor protein can result in activation or inhibition of the protein enzyme. Two main approaches are the most popular of the different molecular docking strategies. The first strategy uses a matching technique that explains protein and ligand as complementary surfaces. The second approach, however, simulates the actual docking process, in which the ligand–protein interaction energies are calculated. Molecular docking plays an important role in the rational drug designing. For a study of interaction of ligand (inhibitor and cofactor) and protein target one can use HEX, BIOSOLVEIT, DOCKING SERVER and other servers listed in Table 12.6.

Plant–Pathogen Interactions

Many microbes establish wide range of interactions with host plants. Some of these are pathogenic and some are symbiotic in nature. Such interactions involve complex recognition events between the plant and the microbe, leading to a cascade of signalling events and regulation of a number of genes is required for, or associated with, the interaction. The combined components of the transcriptomes of both plant and microorganism that are expressed during the interaction give rise to the term "interaction transcriptome". High-throughput methods to study differential gene transcription, or proteomics coupled with bioinformatics will accelerate our understanding of the molecular bases of plant–microbe interactions (Birch and Kamoun 2000; Samac and Graham 2007). For example, Soria-Guerra et al. (2010) conducted a transcriptome profiling study for soybean rust (*Phakopsora pachyrhizi*) to identify soybean rust resistance genes in *Glycine tomentella*. Among 38,400 genes

Table 12.6 List of servers related to inhibitor, cofactor and protein docking

Server	Description/function	URL
SwissDock	Predicts the molecular interactions between a target protein and a small molecule	http://swissdock.vital-it.ch/
DockingServer	Molecular docking from ligand and protein set-up	http://www.dockingserver.com/web
Blaster	Docking program developed by Pharmaceutical Chemistry Department at the California University	http://blaster.docking.og/
Docking At UTMB	Structure-based virtual screening with AutoDock Vina	http://docking.utmb.edu/
Pardock	Fully automated, all-atom energy based ligand docking	http://www.scfbio-iitd.res.in/dock/pardock.jsp
PPDock	Portal Patch Dock is a web server that can be used to dock drugs to the target proteins	http://140.112.135.49/ppdock/
iScreen	Docking and screening the small molecular database on traditional Chinese medicine (TCM) using the LEA3D genetic algorithm	http://iscreen.cmu.edu.tw/
TarFisDock	It docks small molecules into the protein targets in Potential Drug Target Database, and ranks them by the energy score, including their binding conformations	http://www.dddc.ac.cn/tarfisdock/
PLATINUM	Calculates match or mismatch in receptor–ligand complexes and hydrophobic properties of molecules	http://model.nmr.ru/platinum/

monitored using a soybean microarray, 1,342 genes exhibited significant differential expression between uninfected and *P. pachyrhizi*-infected leaves at 12, 24, 48, and 72 h post-inoculation (hpi) in both rust-susceptible and rust-resistant genotypes. Differentially expressed genes were grouped into 12 functional categories, and a large numbers of these genes relate to the basic plant metabolism. These findings provided a better insight into the mechanisms underlying resistance and general activation of plant defense mechanisms in response to rust infection in soybean.

Further, sequencing of EST libraries from pathogen-inoculated or elicitor-treated plants and microarray transcript analyses have enabled the elucidation of genome-wide gene expression changes associated with defence (Ameline-Torregrosa et al. 2006). Samac et al. (2011) used microarray analysis to identify the genes associated with disease defence responses in *M. truncatula*. They compared the genes expressed in response to three pathogens (*Colletotrichum trifolii*, *Erysiphe pisi* and *Phytophthora medicaginis*) and identified genes unique to an interaction.

Fusarium wilt, the most serious disease of pigeonpea, is a common vascular wilt fungal disease caused by *Fusarium* sp. A release draft genome assembly of six strains of different *Fusarium* sp. (Rep and Kistler 2010) gives opportunities to understand the host–pathogen interaction at computational level. In this context, bioinformatics approaches help in understanding the host–pathogen interaction at protein level, in which protein–protein interactions are used to investigate the biological process. Protein–protein interactions are interactions between two or more

Fig. 12.13 An automated protein–protein interaction server

proteins that bind together to carry out their biological function. Protein-protein docking will help understand protein–protein interactions at computational level. HEX, Z-DOCK and other tools are commonly used for protein–protein interaction studies (Fig. 12.13).

Bioinformatics in Molecular Marker Development

For trait analysis using association mapping approaches, and for various other studies on populations including pattern of evolution, population structure, genetic diversity a number of software are available in public domain (Table 12.7). Bioinformatics plays very important role in molecular marker developments, for which several bioinformatics tools and servers are available (Table 12.8). Best optimized primers are essential for good specificity and efficiency. Anyone can design the primer pairs using genomics, mRNA, cDNA, SNP-based sequences. One can design degenerate, expression and universal primers using bioinformatics tools based on servers listed in Table 12.8. For example, Jayashree et al. (2006) have developed a database for EST based simple sequence repeats from cereals and

Table 12.7 Statistical analysis tool and software details with uniform resource locator

Tools/software's name	Description/function	URL (uniform resource locator)
TASSEL	Trait Analysis by Association, Evolution and Linkage; implements general linear model and mixed linear model approaches for association mapping; takes into account population and family structures	http://www.maizegenetics.net/
STRUCTURE	A software package uses multi-locus genotype data to investigate population structure to infer the presence of distinct populations; assigns individuals to populations, detects hybrid zones, identifies migrants and admixed individuals	http://pritch.bsd.uchicago.edu/structure.html
SPAGeDi (Spatial Pattern Analysis of Genetic Diversity)	A computer package primarily designed to characterize the spatial genetic structure of mapped individuals and/or mapped populations using genotype data of any ploidy level	http://ebe.ulb.ac.be/ebe/Software.html
EIGENSTRAT	Uses principal components analysis to explicitly model ancestry differences between cases and controls along continuous axes of variation; the resulting correction is specific to a candidate marker's variation in frequency across ancestral populations; minimizes spurious associations and maximizes power to detect true associations	http://genepath.med.harvard.edu/~reich/Software.htm
MTDFREML	Multiple Trait Diversity Analysis and analysis of variance components	http://aipl.arsusda.gov/curtvt/mtdfreml.html
ASERML	A statistical software package for fitting linear mixed models using restricted maximum likelihood, which is commonly used in plant and animal breeding, and quantitative genetics, and other fields; fits very large and complex data sets efficiently, due to its use of the average information algorithm and sparse matrix methods	http://www.vsni.co.uk/software/asreml
R	A free software environment for statistical computing and graphics; provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering etc.) and graphical techniques, and is highly extendible	http://www.r-project.org/
LDMAP	A program for constructing linkage disequilibrium (LD) maps	http://cedar.genetics.soton.ac.uk/pub/PROGRAMS/LDMAP
SAS	Standard statistical package for traditional statistical analysis	http://www.sas.com/software/sas9/
SPSS	Data mining, statistical analysis and data management softwares	http://www.spss.co.in
NTSys	Discovers patterns and structures in multivariate data	http://www.exetersoftware.com
SigmaPlot	Scientific data and graphing software	http://www.sigmaplot.com

Table 12.8 List of servers used in molecular marker development

Tool/servers name	Description/function	Designated website
Primer3	Widely used program for designing PCR primers	http://frodo.wi.mit.edu/
Gene Fisher	Primer designing based on multiple sequence alignment	http://bibiserv.techfak.uni-bielefeld.de/genefisher/
Web Primer	PCR primer design	http://www.yeastgenome.org/cgi-bin/web-primer
CODEHOP	COnsensus-DEgenerate Hybrid Oligonucleotide Primer	http://blocks.fhcrc.org/codehop.html
PCR Designer	PCR Designer for Restriction Analysis of Sequence Mutations	http://cedar.genetics.soton.ac.uk/public_html/primer.html
Primo Multiplex 3.4	Multiplex PCR Primer Design	http://www.changbioscience.com/primo/primoml.html
Primer Quest	PCR Primers with Probe	http://eu.idtdna.com/scitools/applications/primerquest/
Primo Pro 3.4	PCR Primer Design	http://www.changbioscience.com/primo/primo.html
Primo Degenerate 3.4	Degenerate PCR Primer Design	http://www.changbioscience.com/primo/primod.html
MethPrimer	Design Primers for Methylation PCRs	http://www.urogene.org/methprimer/index1.html
Primaclade	Identifies a set of PCR primers that will bind across the alignment	http://www.umsl.edu/services/kellogg/primaclade.html
Primer3Plus	Pick primers from a DNA sequence	http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi
PrimerBLAST	Finding primers specific to PCR template	http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi
SNP Primers	Creating primers around SNPs in genomic DNA	http://pcrsuite.cse.ucsc.edu/SNP_Primers.html
SSRLocator	Simple Sequence Repeat based primer designing	http://www.ufpel.tche.br/faem/fitotecnia/fitomelhoramento/faleconosco.html
MISA	MIcroSAtellite identification based primer designing	http://pgrc.ipk-gatersleben.de/misa/

legumes. Based on the available resources any one can design EST SSR-based markers for wet-lab experimentation. Large-scale transcriptome assembly using next generation sequencing technologies such as, Roche/454 and Illumina/Solexa, are now used for development of molecular markers, which will serve as a useful resource to accelerate genetic research and breeding applications in legumes. For example, Hiremath et al. (2011) developed 728 SSR, 495 SNP, 387 conserved orthologous sequence (COS) markers, and 2,088 intron-spanning region (ISR) markers in chickpea. Kudapa et al. (2012) predicted for 6,284 intron spanning regions (ISR) covering all the 11 pigeonpea linkage groups.

Mishra et al. (2012) retrieved a total of 18,552 EST sequences (equivalent to 11.3 MB) from the EST database available in the NCBI public domain and analysed for repeat patterns using the tandem repeat finder program at http://c3.biomath.

mssm.edu/trf.html, followed by their assembly using the CAP3 software program (Huang and Madan 1999). After pre-processing, they identified SSR-containing sequences by a perl script-based program, MISA software (MICROSATELLITE identification tool, http://pgrc.ipk-gatersleben.de/misa/). They detected 10,800 unigenes from 18,522 pea EST sequences and screening of 10,800 unigenes by MISA revealed 2,612 (14.1 %) eSSRs in 2,395 (12.9 %) SSR-containing ESTs, from which 577 (24.1 %) primer pairs were designed. Out of these, 68 randomly selected primer pairs showed high rate (48–85 %) of transferability in leguminous species with high level of polymorphism, reproducibility and presence of 3.8 alleles/locus. Similarly, De Caire et al. (2012) retrieved a total of 6,327 mRNA sequences and screened them through a JAVA based programme to design gene-based SSR markers. They successfully identified 45 new polymorphic eSSR markers. e-SSRs identified in these two studies will be used in linkage mapping analyses and provide a good scaffold for comparative mapping in pea and other sequenced legumes.

The molecular markers can be used for linkage mapping using mapping populations developed from biparental crosses. Software like MAPMAKER, QTL-ALL, QTLNETWORK, QUANTO, QU-GENE, QUTIE etc. are used for mapping of markers and oligogenes, while QTL cartographer, QGENE, QTL CAFE, QTL EXPRESS etc. are available for mapping of quantitative trait loci (QTLs). The genes/QTLs detected for target traits need to be confirmed in other replicate studies. Further the marker found linked to the genes/QTLs have to be validated in unrelated germpasm/materials before they can be used for markers-assisted selection (MAS) in plant breeding programmes. Alternatively, marker trait associations can be detected by linkage disequilibrium (LD) based association mapping that uses germplasm collections/breeding lines in the place of biparental mapping populations.

Conclusion and Perspectives

Omics era in the twenty-first century provides us opportunities to understand the legume genome at sequence-structural-functional levels. While legume omics is still in its infancy, it holds great promise, and is expected to yield insights into many aspects of evolution and regulatory mechanisms of legume species. The rapid development of various molecular tools and techniques including large scale analysis of genome organization, gene expression, protein–protein interaction and protein–ligand interaction etc. are generating enormous amount of data, which need to be analyzed and interpreted to develop a biologically meaningful concepts. The need for handling such large amounts of data as forced rapid development of bioinformatics techniques to create, manage and utilize databases of biological information and development of tools and software packages to make efficient and meaningful use of these tools and databases. A variety of software packages are now available to serve various needs of the researchers. However, there is need to develop user friendly bioinformatics tools to decipher functional features of legume genome sequences.

References

Ameline-Torregrosa C et al (2006) Transcriptomic approaches to unravel plant–pathogen interactions in legumes. Euphytica 147: 25–36

Bertioli DJ, Moretzsohn MC, Madsen LH, Sandal N, Leal-Bertioli SC, Guimarães PM, Hougaard BK, Fredslund J, Schauser L, Nielsen AM, Sato S, Tabata S, Cannon SB, Stougaard J (2009) An analysis of synteny of *Arachis* with *Lotus* and *Medicago* sheds new light on the structure, stability and evolution of legume genomes. BMC Genomics 23: 45

Birch PRJ, Kamoun S (2000) Studying interaction transcriptome: coordinated analyses of gene expression during plant–microorganism interactions. In: New technologies for life sciences: a trends guide. Elsevier, London, UK, pp 77–82

Bordat A, Savois V, Nicolas M, Salse J, Chauveau A, Bourgeois M, Potier J, Houtin H, Rond C, Murat F, Marget P, Aubert G, Burstin J (2011) Translational genomics in legumes allowed placing *in silico* 5460 unigenes on the pea functional map and identified candidate genes in *Pisum sativum* L. *G3 (Bethesda)*. Genes Genomes Genet 1(2): 93–103

Cannon SB, May GD, Jackson SA (2009) Three sequenced legume genomes and many crop species: rich opportunities for translational genomics. Plant Physiol 151: 970–977

David P, Sévignac M, Thareau V, Catillon Y, Kami J, Gepts P, Langin T, Geffroy V (2008) BAC end sequences corresponding to the B4 resistance gene cluster in common bean: a resource for markers and synteny analyses. Mol Genet Genomics 280:521–533

De Caire J, Coyne CJ, Brumett S, Shultz JL (2012) Additional pea EST-SSR markers for comparative mapping in pea (*Pisum sativum* L.). Plant Breed 131:222–226

Doyle JJ, Luckow MA (2003) The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. Plant Physiol 131(3):900–910

Dutt S, Singh VK, Marla SS, Kumar A (2010) *In silico* analysis of sequential, structural and functional diversity of wheat cystatins and its implication in plant defense. Genomics Proteomics Bioinformatics 8: 42–56

Garg R, Patel RK, Jhanwar S, Priya P, Bhattacharjee A, Yadav G, Bhatia S, Chattopadhyay D, Tyagi AK, Jain M (2011) Gene discovery and tissue-specific transcriptome analysis in chickpea with massively parallel pyrosequencing and web resource development. Plant Physiol 156(4):1661–1678

Gepts P, Beavis WD, Brummer EC, Shoemaker RC, Stalker HT, Weeden NF, Young ND (2005) Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through genomics conference. Plant Physiol 137:1228–1235

Goffard N, Weiller G (2006) Extending MapMan: application to legume genome arrays. Bioinformatics J 22: 2958–2959

Hiremath PJ, Farmer A, Cannon SB, Woodward J, Kudapa H, Tuteja R et al (2011) Large-scale transcriptome analysis in chickpea (*Cicer arietinum* L.), an orphan legume crop of the semi-arid tropics of Asia and Africa. Plant Biotechnol J 9:922–931

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9: 868-877

Jayashree B, Punna R, Prasad P, Bantte K, Hash CT, Chandra S, Hoisington DA, Varshney RK (2006) A database of simple sequence repeats from cereal and legume expressed sequence tags mined in silico: survey and evaluation. In Silico Biol 6:607–620

Kaló P, Seres A, Taylor SA, Jakab J, Kevei Z, Kereszt A, Endre G, Ellis TH, Kiss GB (2004) Comparative mapping between *Medicago sativa* and *Pisum sativum*. Mol Genet Genomics 272:235–246

Kawaji H, Hayashizaki Y (2008) Genome annotation. Methods Mol Biol 452:125–139

Kudapa H, Bharti AK, Cannon SB, Farmer AD, Mulaosmanovic B, Kramer R et al (2012) A comprehensive transcriptome assembly of pigeonpea (*Cajanus cajan* L.) using Sanger and second-generation sequencing platforms. Mol Plant 5:1020–1028

Kumari A, Singh VK, Fitter J, Polen T, Kayastha AM (2010) Alpha-amylase from germinating soybean (Glycine max) seeds—purification, characterization and sequential similarity of conserved and catalytic amino acid residues. Phytochemistry 71:1657–1666

Kushwaha H, Gupta N, Singh VK, Kumar A, Yadav D (2008) In silico analysis of PCR amplified DOF (DNA binding with one finger) transcription factor domain and cloned genes from cereals and millets. Online J Bioinformatics 9:130–143

Kushwaha H, Gupta S, Singh VK, Rastogi S, Yadav D (2011) Genome wide identification of Dof transcription factor gene family in sorghum and its comparative phylogenetic analysis with rice and *Arabidopsis*. Mol Biol Rep 38: 5037–5053

Kushwaha H, Gupta S, Singh VK, Bisht NC, Sarangi BK, Yadav D (2012) Cloning, in silico characterization and prediction of three dimensional structure of Sbdof1, Sbdof19, Sbdof23 and Sbdof24 proteins from Sorghum [*Sorghum bicolor* (L.) Moench]. Mol Biotechnol 1:12

Lee JM, Grant D, Vallejos CE, Shoemaker RC (2001) Genome organization in dicots. II. Arabidopsis as a "bridging species" to resolve genome evolution events among legumes. Theor Appl Genet 103:765–773

Marla SS, Singh VK (2012) LOX genes in blast fungus (*Magnaporthe grisea*) resistance in rice. Funct Integr Genomics 12: 265–275

Mishra RK, Gangadhar BH, Nookaraju A, Kumar S, Park SW (2012) Development of EST-derived SSR markers in pea (*Pisum sativum*) and their potential utility for genetic mapping and transferability. Plant Breed 131:118–124

Ramírez M, Graham MA, Blanco-López L, Silvente S, Medrano-Soto A, Blair MW et al (2005) Sequencing and analysis of common bean ESTs. Building a foundation for functional genomics. Plant Physiol 137:1211–1227

Rep M, Kistler HC (2010) The genomic organization of plant pathogenicity in *Fusarium* species. Curr Opin Plant Biol 13: 420–426

Samac DA, Graham MA (2007) Recent advances in legume–microbe interactions: recognition, defense response, and symbiosis from a genomic perspective. Plant Physiol 144:582–587

Samac DA et al (2011) Expression of coordinately regulated defence response genes and analysis of their role in disease resistance in *Medicago truncatula*. Mol Plant Pathol 12:786–798

Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 94:441–448

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74:5463–5467

Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K et al (2008) Genome structure of the legume, *Lotus japonicus*. DNA Res 15:227–239

Schmutz J, Cannon JB, Schlueter J, Ma J, Mitros T et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Singh VK, Singh AK, Chand R, Kushwaha C (2011) Role of bioinformatics in agriculture and sustainable development. Intern J Bioinformatics Res 3: 221–226

Singh NK, Gupta DK, Jayaswal PK, Mahato AK, Dutta S, Singh S, Bhutani S et al (2012) The first draft of the pigeonpea genome sequence. J Plant Biochem Biotechnol 21:98–112

Soria-Guerra RE et al (2010) Transcriptome analysis of resistant and susceptible genotypes of Glycine tomentella during Phakopsora pachyrhizi infection reveals novel rust resistance genes. Theor Appl Genet 120:1315–1333

Thornton JW, DeSalle R (2000) Gene family evolution and homology: genomics meets phylogenetics. Annu Rev Genomics Hum Genet 1: 41–73

Udvardi MK (2002) Legume genomes and discoveries in symbiosis research. Genome Biol 3:reports 4028

Varshney RK, May GD (2012) Next-generation sequencing technologies: opportunities and obligations in plant genomics. Brief Funct Genomics 11:1–2

Varshney RK, Close TJ, Singh NK, Hoisington DA, Cook DR (2009) Orphan legume crops enter the genomics era! Curr Opin Plant Biol 12:202–210

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MT, Azam S et al (2011) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotechnol 30:83–89

Varshney RK, Chi S, Saxena RK, Azam S, Sheng Y, Shapre AG, Steven C, Jongmin B, Rosen BD et al (2013) Draft genome sequence of chickpea provides a resource for trait improvement. Nat Biotechnol 31:240–246

Wojciechowski MF, Lavin M, Sanderson MJ (2004) A phylogeny of legumes (Leguminosae) based on analyses of the plastid matK gene resolves many well-supported subclades within the family. Am J Bot 91:1846–1862

Yadav D, Singh VK, Singh NK (2007) *In silico cis*-regulatory elements analysis of seed storage protein promoters cloned from different cultivars of wheat, rice and oat. Online J Bioinformatics 8(2):1–9

Young ND, Udvardi M (2009) Translating *Medicago truncatula* genomics to crop legumes. Curr Opin Plant Biol 12:193–201

Young ND, Cannon SB, Sato S, Kim D, Cook DR et al (2005) Sequencing the gene spaces of *Medicago truncatula* and *Lotus japonicus*. Plant Physiol 137:1174–1181

Zhu H, Choi HK, Cook DR, Shoemaker RC (2005) Bridging model and crop legumes through comparative genomics. Plant Physiol 137:1189–1196

Chapter 13
Genetics and Genomics of Resistance to Rust and Stemphylium Blight in Lentil

Gopesh C. Saha and Fred J. Muehlbauer

Abstract Lentil (*Lens culinaris* Medikus) is an important pulse crop species which grows primarily in South-east Asia, Canada, North America, Middle Eastern countries, and Australia. Lentil crop, like any other food legumes affected by various biotic stresses. In many lentil growing regions, predominance of rust and stemphylium blight are reported causing high yield losses. Rust caused by *Uromyces viciafabae* Pers., an obligately biotroph and stemphylium blight caused by *Stemphylium botryosum* Wallr., a saprophyte. In this review, basic information regarding these two diseases along with inheritance of resistance genes and ongoing molecular breeding efforts to breed resistant lentil genotypes are briefly discussed.

Keywords Lentil rust • Stemphylium blight • *Uromyces vicia-fabae* • *Stemphylium botryosum* • Inheritance • Disease resistance breeding • Linked molecular markers

Introduction

Lentil (*Lens culinaris* Medikus subsp. culinaris) is a diploid (2n = 2x = 14 chromosomes) self-pollinating annual species with a haploid genome size of an estimated 4,063 Mbp (Arumuganathan and Earle 1991). It is an important legume crop and an important source of dietary protein in human diets and animal feed throughout West Asia and North Africa, the Indian subcontinent, North America, South America and Australia (Webb and Hawtin 1981; Erskine 1997). World production of lentil is

G.C. Saha, Ph.D. (✉)
Department of Crop and Soil Sciences, Washington State University, 301, Johnson Hall, Pullman, WA 99163, USA
e-mail: gopesh_saha@wsu.edu

F.J. Muehlbauer, Ph.D.
Grain Legume Genetics Physiology Research Unit, USDA, WSU, Pullman, WA, USA

S. Gupta et al. (eds.), *Legumes in the Omic Era*, DOI 10.1007/978-1-4614-8370-0_13,

estimated at 4.4 million metric tons from an estimated 4.2 million hectares with an average yield of 1,058 kg/ha (FAOSTAT 2011). It is an important component of crop diversification in predominantly cereal based cropping systems in South Asia and also an important legume rotational crop in the US Pacific Northwest. The lentil crop can improve soil nutrient status through symbiotic nitrogen fixation, conserving soil moisture and limiting soil erosion (Muehlbauer et al. 1992). Lentil stubbles that were left standing overwinter and trap snow and reduce the rate of evaporation of soil moisture in spring and prevent erosion (Anonymous 2008). Moisture conservation is important to soil conservation because the additional moisture improves crop growth in the following year. Numerous factors limit yield and seed quality of lentil, including limited moisture availability, salinity, weeds, insect pests and various diseases. Diseases are major factors that limit yields and cause yield instability. Rust caused by *Uromyces vicia-fabae* Pers., an obligately biotroph and stemphylium blight caused by *Stemphylium botryosum* Wallr., a common saprophyte, are the two major diseases of lentil in South Asia and North and East Africa, where lentil is considered nutritional security crop. These two diseases cause variable degrees of damage depending upon the time of its onset and environmental conditions.

Rust Disease

Lentil rust infects a narrow range of living plant hosts and causes substantial yield losses annually in Bangladesh, India, Ethiopia, Morocco and Pakistan. High humidity, cloudy or drizzly weather with temperatures 20–22 °C favors disease development. The disease appears during the flowering and early podding stages in areas with dense crop canopies. Lentil seed contaminated with pieces of rust infected debris with teliospores are the sources of primary infection (Khare 1981; Agarwal et al. 1993). In severe infections the leaves are shed and plants dry prematurely without producing seed or by having shriveled seed. Early infection accompanied by favorable environmental conditions can result in complete crop failure and huge economic losses. Yield losses of 30–70 % have been reported in some years in research plots and 100 % yield loss reported by Negussie et al. (1998). Lentil rust disease was first reported in Bangladesh in year 1974 and caused serious damage.

Taxonomy and Morphology of *Uromyces vicia-fabae*

"The causative pathogen of lentil rust is *Uromyces vicia-fabae* (Pers.) Schroet and is a member of the *Pucciniaceae* and order *Uredinales*. It is an autoecious fungus that completes its life cycle on lentil. Spermagonia are sub-epidermal and globoid. Aecia are sub-epidermal in origin, erumpent later. Aeciospores are elliptical, yellowish brown, measuring 14–22 μm in diameter and with a finely warty wall. Uredia are first sub-epidermal, then erumpent. Urediospores are borne singly on pedicels, mostly echinuate, with three to four germination pores and measure 22–28 μm × 19–22 μm.

Telia are sub-epidermal with origin, then erumpent on leaves but remain covered by the epidermis on stems for an extended period. Teliospores are borne singly on pedicels, are globose to sub-globose, very rarely ellipsoid or ovoid, one celled, measuring 25–40 μm × 18–26 μm, with a single germination pore; the wall is obviously pigmented" (Bayaa and Erskine 1998; Viennot-Bourgin 1949).

Pathogenic Race of Rust Fungus

Singh and Sokhi (1980) identified six pathotypes of rust on the basis of their differential reactions on different cultivars of lentil, pea and sweet pea. Singh et al. (1995) have reported five races, and Conner and Bernier (1982) detected 11 races of *U. viciae-fabae*. Conner and Bernier (1982) speculated that *Vicia*, *Lathyrus* and *Pisum* could be another important source of inoculum and perhaps pathogenic variants due to race specific resistance and selection pressure on the pathogen.

Germplasm Sources for Rust Resistance

ICARDA developed 90 lentil lines resistant to rust, and some of them have combined resistance against a range of biotic and abiotic stresses (Sarker et al. 2002). For breaking the "bottleneck" of narrow genetic base of lentil in South Asia and combating two major diseases of lentil, rust and stemphylium, ICARDA scientists are working with Bangladesh counterparts to introgress desirable genes to improve Bangladeshi land races. These Bangladeshi land races belong to the glex pilosae as described by Barulina (1930) as one of the microsperma types. This approach will help to improve resistance to pulse diseases in Bangladesh (ICARDA 2004). Negussie et al. (2005) reported four cultivars "Gudo", R-186, FLIP-87-66L and FLIP-89-60L with different levels of rust resistance. Based on final rust severity, area under disease progress curve (AUDPC), area under the pustule density curve (APDC) and apparent infection rate (rG) values, Gudo and R-186 were grouped as having a high level of resistance, FLIP-89-60L as moderately susceptible and FLIP-87-66L as intermediate between susceptible and moderately susceptible lines

Disease Rating and Assessment for Rust

Khare et al. (1993) developed a 9-point disease severity scale. The scale was then categorized according to Singh and Sandhu (1988) as: 1 = resistant (no infection), 3 = moderately resistant (10 % leaf area infected), 5 = moderately susceptible (10.1–25 % leaf area infected), 7 = susceptible (25.1–50 % leaf area with stems also infected), 9 = highly susceptible (>50 % leaf area with stem and pods heavily infected). Chen (2007) modified the Khare et al. (1993) scale of disease severity

based on field conditions, where 1 = 0–10 % leaf area infected, 3 = 11–30 % leaf area infected, 5 = 31–50 % leaf area infected, 7 = 51–70 % leaf area infected, 9 = More than 70 % leaf area infected.

Genetics and Genomics of Rust Resistance

The rust pathogen requires the induction of a subset of fungal genes essential for infection. The infection mechanisms are sophisticated, and include the ability to detect stomata, the entry portal for many rust fungi, and to suppress host resistance responses. During the first hour of infection, the fis1 protein is localized exclusively in leaf mesophyll cells closely surrounding the rust infection site. As a result, a hypersensitive response is expressed by the resistant host after the first hour (Ayliffe et al. 2002).

Most studies on genetics of rust resistance in lentil have revealed that resistance is monogenic and dominant (Sinha and Yadav 1989; Singh and Singh 1992). Kumar et al. (1997) reported that resistance to *Uromyces fabae* in five lentil genotypes (L 2991, L 2981, L 1534, L 178 and HPLC 8868) was governed by single dominant genes; whereas in one genotype, Precoz, it was conditioned by two dominant genes. Chahota et al. (2002) reported that resistance to rust is controlled by duplicate, non-allelic and non-linked dominant genes. Negussie et al. (2005) has not ruled out the likelihood of monogenic resistance based on their research result. Rust resistance in lentil is controlled by three genes and two of the genes (*Urf1* and *Urf2*) were dominant in nature (Basandrai et al. 2007). Saha et al. (2010a) identified one sequence related amplified polymorphic marker (SRAP), F7XEM4a, closely linked to rust resistance in lentil. This marker is located 7.9 cM from the resistance gene on our linkage group 3 and suggested that F7XEM4a marker could be used for marker-assisted selection for resistance after validation. Gupta et al. (2012) used interspecific F_2 population to map rust resistance but felt the need of more markers to saturate the linkage map and made an attempt to use cross genera SSR markers due to lack of polymorphisms in lentil genome.

Continuous cultivation of varieties with race specific resistance in large areas increases selection pressure on the pathogen that may lead to the formation of new races capable of infecting the previously resistant varieties. There are several strategies for developing varieties with durable resistance. These include multilines (Marshall 1977), partial resistance/slow rusting (Wilcoxson et al. 1975) and gene pyramiding (Green 1975; Pederson and Leath 1988).

Stemphylium Blight Disease

Stemphylium blight is a serious threat to lentil (*Lens culinaris* Medik.) cultivation in some parts of the world, especially in South Asia including Bangladesh, Northeast India and Nepal. This disease has also been reported in lentil fields in Egypt, Syria

and North America (Erskine and Sarker 1997; Bayaa and Erskine 1998; ICARDA 2004; Banniza 2005). Stemphylium blight disease starts as pinhead-sized light brown or colored spots on leaflets of plants in dense populations. The spots enlarge rapidly and within 2–3 days they cover the entire leaflet resulting in defoliation and death of young plants. In severe cases the crop may exhibit a blighted appearance causing large-scale defoliation; however, the pods may remain green. In South Asia, temperatures of 18–20 °C and relative humidity of over 85 % have been reported to favor the development of disease (Erskine and Sarker 1997). It has also been reported in Canada that *S. botryosum* prefers temperatures above 25 °C, 85 % relative humidity and a minimum of 8 h leaf wetness (Northover and Dokken 2009). The disease results in yield losses that exceed 60 % in severely infected fields. Disease intensity as high as 83 % was observed on an unsprayed local susceptible lentil cultivar in Bihar state of India, causing more than 90 % yield loss (Sinha and Singh 1993). The diverse host range of *Stemphylium botryosum*, which includes leguminous and non-leguminous species in different parts of the world, demonstrates its adaptability to different genotypes and environments (du Toit and Derie 2001). Stemphylium blight was identified in farmer and research fields by Sen and Das (1964), Nene et al. (1984) in India, Kaiser (1972) in Iran and Bakr and Zahid (1986) in Bangladesh, and Simay (1990) in Hungary. Relatively severe and widespread distribution of disease was reported in 2007 at Sasketchewan, Canada. With the increase of lentil production and deployment of resistance to ascochyta blight and anthracnose in new cultivars, stemphylium blight has become a more serious problem (Vandenberg and Morrall 2002).

Taxonomy and Morphology of *Stemphylium botryosum*

The asexual stage of the causal organism of stemphylium blight is *Stemphylium botryosum* Wallr.; whereas, *Pleospora herbarum* is the sexual stage. The fungus is commonly referred as anamorph. Stemphylium blight is a ubiquitous, dematiaceous filamentous fungus that belongs to the kingdom Fungi, phylum Ascomycota, class Ascomycetes, order Pleosporales, family Pleosporacea (Inderbitzin et al. 2009). In medical science, the fungus is considered an allergen (Larone 2002).

Morphological and developmental characters such as size and shape of the conidia, conidiophores, ascospores and the size and time of maturation of pseudothecia were useful for diagnosing species (Câmara et al. 2002). "Conidiophores are short, arise singly or in groups and are aseptate and swollen at the apex. After a conidium is produced, the end of the conidiophore grows out and produces a new cell and a new conidium. The conidiophore may grow to a considerable length and have a nodulose appearance. Conidia are olive brown, muriform and echinulate measuring 24–40 μm × 14–25 μm. Conidia are oblong with three to four septae and often constricted at the center by medium cross walls. Perithecia are globose, membranous and black, and sometimes have a slender neck. Asci (183–267 μm × 27–37 μm) are oblong to clavate with outer and inner walls. Ascospores (32–48 μm × 12–21 μm) are elongate to ovate, characteristically with seven cross walls and three to five

longitudinal septa, and yellowish to brown in color and muriform when mature" (Bayaa and Erskine 1998). Estimates of the numbers of described Stemphylium species vary from around 20 to 30 (Câmara et al. 2002; Kirk et al. 2001) to up to 150 (Wang and Zhang 2006). The phylogenetic relationships of 43 isolates representing 16 species of Stemphylium were inferred from ITS and glyceraldehyde-3-phosphate dehydrogenase (gpd) gene sequence data (Marcos et al. 2002).

Resistant Germplasm Sources

It has been reported that Barimasur-4 (developed from a local cultivar of Bangladesh, Utfala) shows significant resistance against rust and stemphylium blight (Erskine and Sarker 1997). Preliminary screening at the Crop Development Center (CDC) of University of Saskatchewan showed that "Crimson", "Eston" and ILL 4605-2 and ILL-8008 have good resistance against stemphylium blight. Podder et al. (2013) reported the highest frequency of resistance to stemphylium blight in *L. lamottei* followed by *L. ervoides* and identified as potential sources for developing new commercial cultivars with multiple or single disease resistance.

Disease Ratings and Assessments and Lab Culture of Inoculum

Horsfall-Barrat's logarithmic scale had unequal intervals in disease scores and is difficult to use for quantitatively inherited traits. To overcome this problem Hashemi et al. (2005) modified this scale to a 0–10 linear semi-quantitative scale. This scale considered disease development pattern consisting of the appearance of chlorotic spots followed by gradual defoliation of plants (0 = free of disease, 1 = a few tiny tan spots, 2 = few small to large chlorotic spots, 3 = expanding lesions on leaves to defoliation started, 4 = 20 % nodes on main stem showing necrotic symptoms and defoliation, 5 = 40 % nodes on main stem showing necrotic symptoms and defoliation, 6 = 60 % nodes on main stem showing necrotic symptoms and defoliation, 7 = 80 % nodes on main stem showing necrotic symptoms and defoliation, 8 = 100 % leaves defoliate but small green tip recovering, 9 = 100 % leaves defoliate but stem still green, 10 = Completely dead). Kumar (2007) used this scale (0–10) for stemphylium blight screening. A disease rating scale from 1 to 5 was used for scoring leaf spots in alfalfa caused by *S. botryosum* (Salter and Leath 1991). Koike et al. (2001) used a sign scale (− = no disease; + = small leaf spot <5 mm; ++ = medium leaf spot) for scoring spinach leaf spot disease caused by *S. botryosum*.

Stemphylium botryosum colonies grow rapidly on a variety of media and mature with in 5 days at 25 °C on potato dextrose agar (Hashemi et al. 2005). On most media, it produces velvety to cottony gray, brown or brownish-black or black colonies (Larone 2002). The production of conidia in abundance under laboratory conditions is difficult, even when it is grown on PDA and or V8 juice agar under

alternate cycle of 12 h light and 12 h darkness (Chowdhury et al. 1996; Mehta 1998). The use of mycelial suspensions in disease screening has been found to be as efficient as spore suspensions (Hashemi et al. 2005).

Genetics of the Stemphylium Blight

There has been limited study on genetics of stemphylium blight in lentils. However, resistant varieties were found to have a thicker cuticle, thicker epidermal cell layer, thicker cortical layers, fewer stomata and a large number of epidermal hairs compared to the susceptible lines (Chowdhury et al. 1997). There has been limited study on genetics of stemphylium blight resistance. Saha et al. (2010b) studied the inheritance of disease resistance and found the presence of dominant genes. They identified significant additive and epistatic gene actions affected the QTLs of resistance. One significant QTL was detected based on disease scores from 1-year data, while three significant QTLs were detected from another year data. The QTL $QLG4_{80–81}$ was common in both years and accounted for 25.2 and 46.0 % of the variation respectively in their respective years. Two SRAP markers, ME5XR10 and ME4XR16c, and one RAPD marker, UBC34, located on linkage group 4, were significantly associated with the $QLG4_{80–81}$ in both crop years. After validation, the more tightly linked ME4XR16c marker may be used for marker-assisted selection for stemphylium blight resistance. The frequency distribution within a lentil RIL population developed from the cross Barimasur-4 × CDC Milestone in both field and controlled environments revealed quantitative inheritance (Kumar 2007). Mihov and Stoyanova (1998) described that lentil cultivars in Bulgaria (Naslada and Stella) also possess complex resistance to stemphylium blight.

Conclusion

For many years, institutes in Bangladesh, Nepal, India, Pakistan, and Ethiopia have put forth considerable effort toward combating rust and stemphylium blight diseases of lentil with minimal progress. After the establishment of International Center for Agricultural Research in the Dry Areas (ICARDA) in 1977, lentil research experienced significant momentum in the developing world. ICARDA played an essential role in collecting and characterizing germplasm from all over the world for resistance to diseases including stemphylium blight and rust. The center now has a collection of 10,000 accessions and 500 related wild species. These materials have been evaluated in collaboration with national institutions for disease resistance and lines with good resistance have been identified. Some of the lines have been released as resistant varieties in their nationalized program or used as a source of resistance in their respective breeding programs. Some material has been released with rust and stemphylium blight resistance. Barimasur-2 was the first

rust resistant variety released in Bangladeh; Bakria (ILL 4605), Bichette (ILL 5562) and Hamira (ILL 6238) released in Morocco; Adaa and Alemaya released in Ethiopia; NIAB Masoor 2006 released in Pakistan have a high level of resistance. Barimasur 3, 4, 5 and 6 are resistant to both rust and stemphylium blight and Bulgarian varieties, Naslada and Stella, have shown significant resistance to stemphylium blight. Understanding the genetics and identification of markers linked to the genes for resistance to rust and stemphylium blight will assist breeding programs toward improving resistance against these two major diseases of lentil. Different marker systems are available for lentil and several breeding programs worldwide are utilizing those markers for gene pyramiding to combat diseases. The emerging field of proteomics and metabolomics of lentil will help to understand the functional components of genomics and plant–pathogen relationships. With the advance knowledge of functional genomics and expression analysis, scientists will better understand the disease resistance mechanisms and will successfully able to clone the potential genes of interest.

References

Agarwal SC, Singh K, Lal SS (1993) Plant protection of lentil in India. In: Erskine W, Saxena MC (eds) Lentil in South Asia. International Center for Agricultural Research in the Dry Area, Aleppo, pp 147–165

Anonymous (2008) Agriculture knowledge centre. Saskatchewan agriculture. http://www.agriculture.gov.sk.ca/Default.aspx?DN=1a2fc0cd-ddb4-47b6-8d29-91dff9dde90a. Accessed 21 June 2013

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol 9:208–218

Ayliffe MA, Roberts JK, Mitchell HJ, Zhang R, Lawrence GJ, Ellis JG, Pryor TJ (2002) A plant gene upregulated at rust infection sites. Plant Physiol 129:169–180

Bakr MA, Zahid I (1986) Stemphylium blights a new foliar disease of lentil in Bangladesh. Bangladesh J Plant Pathol 2(1):69–70

Banniza S (2005) Investigation into the biology of *Stemphylium botryosum*, a potentially new pathogen in lentil production in Saskatchewan. PulsePoint. Saskatchewan Pulse Growers, Sunrise Publ., Saskatoon, SK, p 7

Barulina H (1930) Lentils of the USSR and other counties. Bull Appl Bot Genet Plant Breed 40:225–238

Basandrai D, Basandrai AK, Sarker A, Mishra SK, Thakur HL, Thakur SK (2007) Genetics of rust resistance in lentil (*Lens culinaris* ssp. *culinaris*). J Lentil Res 3:28–31

Bayaa B, Erskine W (1998) Lentil pathology. In: Allen D, Lenné J (eds) Pathology of food and pasture legumes. Commonwealth Agricultural Bureaux International, Andhra Pradesh, pp 423–472

Câmara MPS, O'Neill NR, van Berkum P (2002) Phylogeny of *Stemphylium spp*. based on ITS and glyceraldehyde-3-phosphate dehydrogenase gene sequences. Mycologia 94:660–672

Chahota PK, Gupta VP, Sharma SK (2002) Inheritance of rust resistance in lentil. Indian J Genet Plant Breed 62:226–227

Chen W (2007) Stemphylium blight disease scoring in field condition. Research Geneticist, USDA-ARS, Washington State University, Pullman, WA

Chowdhury AM, Ahmed A, Zaman M, Bakr MA (1996) Sporulation of *Stemphylium botryosum* Wallr. J Mycopathol Res 34:69–71

Chowdhury AM, Ahmed A, Zaman M (1997) Studies on the defence structural factors of some susceptible and resistant varieties of lentil plants. J Mycopathol Res 35:35–39

Conner RL, Bernier CC (1982) Race identification in *Uromyces vicia-fabae*. Phytopathology 72:687–689

du Toit LJ, Derie ML (2001) *Stemphylium botryosum* pathogenic on spinach seed crops in Washington. Plant Dis 85:920

Erskine W (1997) Lessons for breeders from land races of lentil. Euphytica 93:107–112

Erskine W, Sarker A (1997) Lentil: the Bangladesh breakthrough. ICARDA Carvan no. 6. Int. Ctr. for Agric. Res. in Dry Areas, Aleppo

FAOSTAT, database (2011) http://faostat.fao.org/site/291/default.aspx. Accessed 21 June 2013

Green GJ (1975) Virulence changes in *Puccinia graminis f.* sp. *Tritici* in Canada. Can J Bot 53:1377–1386

Gupta M, Verma B, Kumar N, Chahota RK, Rathour R, Sharma SK, Bhatia S, Sharma TR (2012) Construction of intersubspecific molecular genetic map of lentil based on ISSR, RAPD and SSR markers. J Genet 91(3):279–287

Hashemi P, Vandenberg A, Banniza S (2005) Developing a protocol for large scale inoculation of lentil germplasms with *Stemphylium botryosum*. Abstract: In Proceedings of Plant Canada 2005, Edmonton, AB, 15–18 June

ICARDA (2004) Celebration as Bangladesh produces more lentil. ICARDA Caravan 20/21:11–12. Int. Ctr. for Agric. Res. in Dry Areas, Aleppo

Inderbitzin P, Mehta JYR, Berbee ML (2009) Pleospora species with *Stemphylium* anamorphs: a four locus phylogeny resolves new lineages yet does not distinguish among species in the Pleospora herbarum clade. Mycologia 101(3):329–339

Kaiser WJ (1972) Occurrence of three fungal diseases of chickpea in Iran. FAO Plant Prot Bull 20:74–78

Khare MN (1981) Diseases of lentil. In: Webb C, Hawtin GC (eds) Lentils. CAB International, Farnham Royal, UK, pp 163–172

Khare MN, Bayaa B, Beniwal SPS (1993) Selection methods for disease resistance in lentil. In: Singh KB, Saxena MC (eds) Breeding for stress tolerance in cool-season food legumes. Wiley, Chichester, UK, pp 107–121

Kirk PM, Cannon PF, David JC, Stalpers JA (eds) (2001) Ainsworth and Bisby's dictionary of the fungi. CABI Publishing, Wallingford, Oxon, p 655

Koike ST, Henederson DM, Butler EE (2001) Leaf spot disease of spinach in California caused by *Stemphylium botryosum*. Plant Dis 85:126–130

Kumar P (2007) Genetics of resistance to stemphylium leaf blight of lentil (Lens culinaris) in the cross Barimasur-4×CDC Milestone. M.S. Thesis, University of Saskatchewan, Canada

Kumar V, Singh BM, Singh S (1997) Genetics of lentil resistance to rust. Lens Newsl 24(1/2):23–25

Larone DH (2002) Medically important fungi: a guide to identification, 4th edn. American Society for Microbiology, Washington, DC

Marcos PSC, O'Neil NR, van Berkum P (2002) Phylogeny of Stemphylium spp. based on ITS and glyceraldehyde-3-phosphate dehydrogenase gene sequences. Mycologia 94:660–672

Marshall DR (1977) The advantages and hazards of genetic homogeneity. In: Day PR (ed) The genetic basis of epidemics in agriculture. Academy of Sciences, New York, pp 1–20

Mehta YR (1998) Severe outbreak of Stemphylium leaf blight, a new disease of cotton in Brazil. Plant Dis 82:333–336. doi:10.1094/PDIS.1998.82.3.333

Mihov M, Stoyanova M (1998) Results from improving investigation in lentil breeding in Bulgaria. Rasteniev'dni Nauki 35:18–21 (Abstr)

Muehlbauer FJ, Summerfield RJ, Kaiser WJ, Boerboom CM, Welsh-Maddux MM, Short RW (1992) Principles and practice of lentil production. Electronic publication at http://www.ars.usda.gov/is/np/lentis/lentils.htm. Accessed 24 June 2013

Negussie T, Bejiga G, Million E (1998) Lentil rust outbreak in Gimbichu district. AgriTopia, Ethiopia

Negussie T, Pretorius ZA, Bender CM (2005) Component of rust resistance in lentil. Euphytica 142:55–64

Nene YL, Sheila VK, Sharma SB (1984) *Stemphylium sarciniformae*. Wilt. In a world list of chickpea (*Cicer arietinum* L.) and pegionpea (*Cajanus cajan*. L. Millsp.) Pathogen. ICRISAT, Patancheru P.O., Andhra Pradesh, p 502

Northover P, Dokken F (2009) Rising the top: Stemphylium blight of lentils. Crop Prod News 31(8):5–7

Pederson WL, Leath S (1988) Pyrimiding major genes for resistance to maintain residual effects. Annu Rev Phytopathol 26:369–378

Podder R, Banniza S, Vandenberg A (2013) Screening of wild and cultivated lentil germplasm for resistance to stemphylium blight. Plant Genet Resour 11:26–35. doi:10.1017/S1479262112000329

Saha GC, Sarker A, Chen W, Vandemark GJ, Muehlbauer FJ (2010a) Identification of markers associated with genes for rust resistance in *Lens culinaris* Medik. Euphytica 175:261–265

Saha GC, Sarker A, Chen W, Vandemark GJ, Muehlbauer FJ (2010b) Inheritance and linkage map positions of genes conferring resistance to stemphylium blight in lentil. Crop Sci 50:1831–1839

Salter R, Leath K (1991) Spring blackstem and leafspot resistance. In: Fox CC, Berberet R, Gray FA, Grau CR, Jesse DL, Peterson MA (eds) Standard tests to characterize alfalfa cultivars. North American Alfalfa Improvement Conference, Beltsville, MD, p 11

Sarker A, Malhotra RS, Khalil S, Erskine W (2002) Food legume improvement at ICARDA. ICARDA Annual Report, Aleppo

Sen G, Das PK (1964) Effect of vitamins, hormones and trace elements on the growth and sporulation of Stemphylium sarciniformae, causal organism of leaf spot of gram. Plant Dis Rep 48:37–40

Simay EI (1990) Occurence of Epicoccum and Stemphylium leaf spot of Lens culinaris. Lentil Exp News Serv 17(1):28–30

Singh K, Sandhu TS (1988) Screening of cultivars of lentil for resistance to rust. Lentil Exp News Serv 15:28–29

Singh JP, Singh IS (1992) Genetics of rust resistance in lentil (*Lens Culinaris*). Indian J Agric Sci 62:337–338

Singh SJ, Sokhi SS (1980) Pathogenic variability in *Uromyces viciae-fabae*. Plant Dis 64:671–672

Singh K, Rawal HS, Singh S, Gumber RK (1995) Physiological specialization in Uromyces fabae causing rust of lentil. Lentil Exp News Serv 22:46–47

Sinha JN, Singh AP (1993) Effect of environment on the development and spread of stemphylium blight of lentil. Indian Phytopathol 46:252–253

Sinha RP, Yadav BP (1989) Inheritance of resistance to rust in lentil. Lens Newsl 16:41

Vandenberg A, Morrall RAA (2002) Pulse crop variety development strategies in Saskatchewan. Saskatchewan Pulse Growers Pulse Days, Saskatoon, SK

Viennot-Bourgin G (1949) Les champignons parasites des plantes cultivées. Masson and Cie, Paris, France, p 1850

Wang Y, Zhang XG (2006) Three new species of Stemphylium from China. Mycotaxon 96:77–81

Webb C, Hawtin G (1981) Lentils. Commonwealth Bureau, Page Bros, Norwich

Wilcoxson RD, Skovmand B, Atif AA (1975) Evaluation of wheat cultivars for the ability to retard development of stem rust. Ann Appl Biol 80:275–281

Chapter 14
Genomics in Studying the Legume Genome Evolution

Jitendra Kumar, Ekta Srivastava, Mritunjay Singh, and Aditya Pratap

Abstract Molecular markers, genetic and physical maps, ESTs, BAC clones and other genomic resources have been developed in legumes since past few years. Next generation sequencing provided the whole genome sequences of *Medicago truncatula*, *Lotus japonicus*, and *Glycine max*. Further next generation sequencing techniques have rapidly moved the genome scale analysis in less characterized legume species such as chickpea (*Cicer arietinum*), pigeon pea (*Cajanus cajan*), common bean (*Phaseolus vulgaris*) and cowpea (*Vigna unguiculata*); now the launch of third-generation sequencing technologies would further enhance super-scaffolding of genome assemblies into large pseudo molecule. These genomic analysis developments helped to characterize legume family and established macro- and microsyntenic relationships among different taxa of legumes. Analysis of genome sequences of different legumes and use of molecular markers indicated that whole genome duplication has played an important role in evolution of legume genome. These whole genome duplication events have led to evolution of individual genes and gene families in legume species. During the course of evolution of legume genomes, genomic regions responsible for modifying traits from wild to cultivated species (usually known as domestication syndrome) differed variably in terms of rate and the order. In such a situation, genomics encourages to find out the ways of modification in characters, which have been occurring in due course of crop domestication. Thus development in genomic research has revolutionized the knowledge of legume genome evolution.

Keywords Genome evolution • Food legumes • Synteny • Phylogenetic analysis • Whole genome sequencing • Molecular taxonomy

J. Kumar, Ph.D. (✉) • E. Srivastava, M.Sc. • M. Singh, M.Sc. • A. Pratap, Ph.D.
Crop Improvement Division, Indian Institute of Pulses Research, G.T. Road, Kanpur, Uttar Pradesh 208017, India
e-mail: jitendra73@gmail.com

S. Gupta et al. (eds.), *Legumes in the Omic Era*, DOI 10.1007/978-1-4614-8370-0_14,

Introduction

Legumes belong to family Fabaceae which is the third-largest family of flowering plants. These are vitally important to agriculture and the environment. Legumes provide a substantial fraction of all nutritional protein and reduce the need for agricultural chemicals due to their capacity for symbiotic nitrogen fixation. This old and diverse plant family comprises of ~20,000 species (Doyle and Luckow 2003; Gepts et al. 2005). Taxonomically legume family was classified in different subfamily, clades, genera and species on the basis of initial fossil records and morphological features. However, during past three decades, tremendous progress has been made in genomics, which was used to understand the genome organization and evolution of legume family (reviewed in references Young et al. 2003; Zhu et al. 2005). Genomics tools such as whole genome sequences, gene sequences, EST sequences and molecular markers have been exploited to resolve the level of similarity and dissimilarity among various taxa of legumes. Further next generation sequencing has opened up the possibilities of generation of vast amount of genome sequences of non-model legume species at such as pigeonpea and chickpea. Based on available genome sequences of Lotus, *Medicago*, soybean, pigeonpea and chickpea in public domain (Cannon et al. 2009; Schmutz et al. 2010; Varshney et al. 2012, 2013; Sato et al. 2008; Young et al. 2011) have demonstrated broad-scale conservation of genes and gene order (Boutin et al. 1995; Choi et al. 2004b), as well as microsynteny (Cannon et al. 2003; Gualtieri et al. 2002; Mudge et al. 2005; Yan et al. 2004) among different legume taxa. Gene based markers were also used to create an integrated map and to infer genome wide synteny among legume species (Boutin et al. 1995; Choi et al. 2004a; Lee et al. 1999; Menancio-Hautea et al. 1997; Simon and Muehlbauer 1997). Previous studies based on available plant genome sequences in non-legume demonstrated that polyploidy and whole (large-scale)-genome duplication (WGD) play an important role in shaping legume genomes during the course of evolution (*Arabidopsis* Genome Initiative 2000; Yu et al. 2005). Hence these events are important source of understanding evolution of legume genome and domestication of present cultivated species of food legumes. This chapter has thus focused the evolution of legumes genome in the light of genomics developments.

Development of Legume Genomics and Genome Sequencing

During the past years, molecular markers, genetic, physical maps, ESTs, and BAC clones etc. have been developed in legumes (see Kumar et al. 2011 for details). These genomic resources along with next generation sequencing have helped to provide the draft genome sequences of *Medicago truncatula and Lotus japonicus* belonging to galegoid clade and soyabean, chickpea and pigeonpea belonging to millettioid clade. These two clades of sub family Pipilionoidae were diverged ~54 million year ago (Mya) from each other. These species of first clade had a common ancestor ~40 Mya and small genome size. Moreover time of divergences of

these species is also estimated near the time of evolution of most agriculturally important tribes belonging to Papilionoideae. Subsequently the complete, high-quality draft genome sequence of the soybean genome and more recently of pigeonpea have been published (Sato et al. 2008; Varshney et al. 2012, 2013). The next generation sequencing has rapidly moved the genome scale analysis in less characterized legume species such chickpea, common bean and cowpea. Taking the more complete of the genome sequences as a point of reference, we can consider ways in which other legume genomes may be similar or different. In chickpea (*C. arietinum*), next generation led to development of trancriptome, EST and hundreds of different (SNP) and conserved genetic marker sequences used in mapping. Analysis of BAC clone sequences in common bean has led to development of SSR markers, which have provided a basis of integrating the physical and genetic maps of *Phaseolus*. Moreover, high throughput physical mapping by whole genome profiling together with the third-generation sequencing technologies such as of Apacific Biosances will further enhance super-scaffolding of genome assemblies into large pseudo molecules (Young and Bharti 2012).

Genomics in Taxonomic Classification

Characterization of Legume Family

Fossils and phylogenetic records (Schrire et al. 2005a, b) suggest that members of legume family originally evolved in arid and/or semi-arid regions along the Tethys seaway during the early Tertiary period (Herendeen 1992). The West Gondwanan hypothesis also supported a "moist equatorial megathermal" origin of legumes during the mid to late Cretaceous (Raven and Axelrod 1974; Polhill and Raven 1981). Tertiary legume diversification immediately followed the origin of the family. Because legumes are highly diverse in tropical to subtropical Africa and South America, these regions also indicate possible candidates for origin of this family (Pan et al. 2010). Studies suggest a minimum age of 84 million years ago (Mya) for an internal calibration point of the split of Fagales from Cucurbitales, while origin of Fabaceae has been estimated 74–79 Mya (Soltis et al. 2000). The fossil record of the Fabaceae is abundant and diverse, particularly in the Tertiary. Lavin et al. (2005) used the tertiary macrofossils of the Leguminosae as time constraints and molecular data and viewed that the first definitive legumes appeared during the Late Paleocene (~56 Mya) (Herendeen and Wing 2001; Wing et al. 2004). Three oldest clades namely caesalpinioid, mimosoid, and papilionoid were evolved in approximately the same age range of 39–59 Mya. These traditionally recognized subfamilies of legumes and other taxonomically large clades within these subfamilies (genistoids) are recorded from the fossil record soon afterward, beginning around 50–55 Mya (Herendeen 1992). The majority of legume species (approx. 13,800 species) are in the Papilionoideae subfamily and ~3,270 species are in the Mimosoideae subfamily (Lewis et al. 2005). Remaining species are placed in the Caesalpinoideae.

Based on molecular systematics, some caesalpinoid clades are basically placed in the papilionoid subfamily, some along a grade leading to the mimosoid subfamily, and some in separate lineages. Papilionoid subfamily involved most domesticated legume species include the various beans in the millettioid clade; and the peas, vetches (such as faba bean), and clovers in the "hologalegina" clade (also called "galegoid" or "cool-season legume" clade). Based on inverted repeat loss in the chloroplast genome of most angiosperms, galegoid clade have been grouped into "inverted repeat loss clade" (IRLC), and the robinioid clade. The IRLC clade contains *Medicago*, *Cicer* (chickpea and grass pea), *Trifolium* (clovers), *Vicia* (vetches and faba bean), *Pisum* (several pea species), *Glycyrrhiza* (liquorice), and *Lens* (lentil). The robinioid clade includes *Lotus japonicus*, *Lotus tetragonolobus* (asparagus pea), *Sesbania* (a forage and green manure used in flooded rice fields), and *Robinia* (containing the black locust tree, used ornamentally and for durable timber). Nevertheless, a large number of other economically important species are found in the Mimosoideae and early-diverging clades, including a vast number of little-studied species (many of them tropical). Papilionoidae also includes two minor clades.: dalbergioid clade [*Arachis* (peanut) and more than a thousand other species] and genistoid clade [lupin, *Lupinus* sp.] (Lewis et al. 2005; Young and Bharti 2012). A simplified phylogenetic relationship established among different taxa of legume family and their period of evolution has been presented earlier (Cannon et al. 2011).

Syntenic Relationships Among Different Taxa of Legumes

Based on genetic mapping of sequenced based markers and large scale similarity, searches between sequenced genomes were utilized to understand the syntenic relationship among different taxa of legumes (Ahn and Tanksley 1993; Devos and Gale 2000; Gale and Devos 1998). However, a hybrid approach involving comparison of sequenced genetic markers of one species with a species characterized less on the basis of genome sequence using has also been used to study extensive synteny within taxonomical groups as well as among species of within a clade (Bertioli et al. 2009; Muchero et al. 2009; Nayak et al. 2010.). *Medicago* and *Glycine*, which diverged ~55 Mya belonging to two different clades had extensive synteny, which often extends up to whole chromosome arms (Lavin et al. 2005). Though pigeonpea and soybean separated to each other a long period back (~20–30 Mya), these two species of same clade showed high level of synteny (Varshney et al 2012). For study macrosynteny, genetic mapping of common markers or *in silico* mapping of homologous sequences was used while in identification of microsynteny a short, physically defined DNA contig is used.

Macrosynteny

Choi et al. (2004a, b) used genome sequence of *M. truncatula* as a central point of comparison, level of macrosynteny across papilllionoids. For this purpose,

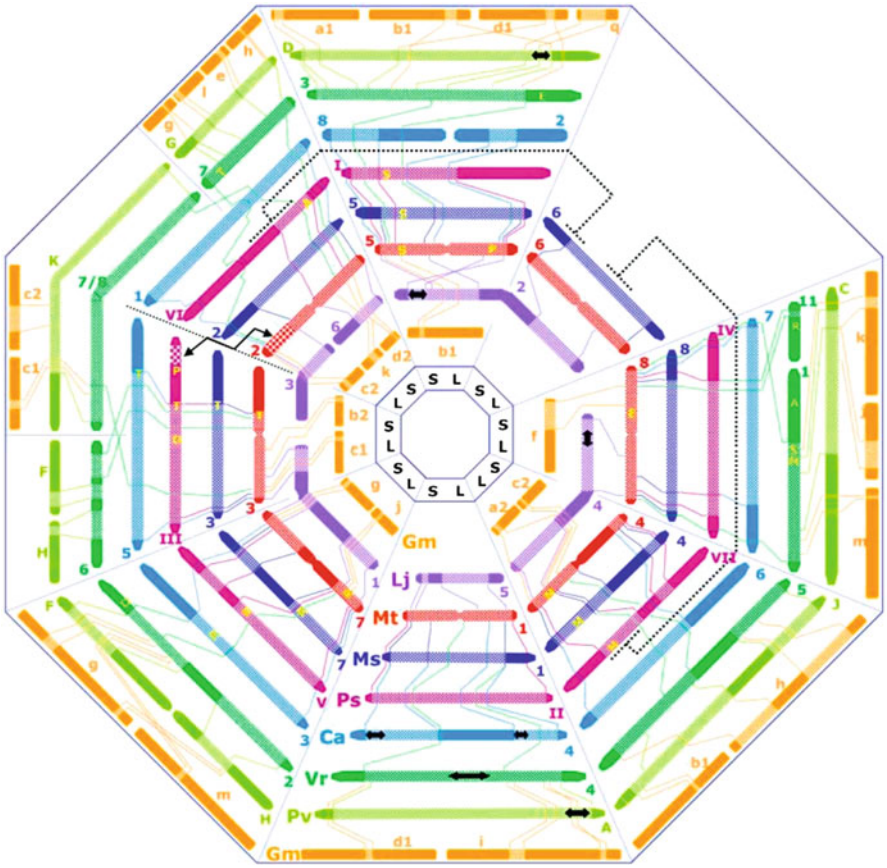


Fig. 14.1 A simplified consensus map for eight legume species [Reprinted from Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, and Cook DR (2004) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci USA 101:15289–15294. With permission from National Academy of Sciences. © (2004) National Academy of Sciences, USA]

molecular markers were developed from *Mt* and *Arabidopsis* sequences and ~50 such putatively orthologous markers were mapped in *M. truncatula*, alfalfa, pea, mungbean and chickpea. In addition to this, 60 additional markers showed homology with genetic markers of soybean were also mapped in *M. truncatula*. Further, 63 pairs of sequenced BAC clones were used to develop the macrosyntenic relationship between *M. truncatula* and *L. japonicus*. These syntenic relationships helped to establish a simplified consensus comparative map of eight legume species, which is well correlated with the phylogenetic distance of these legume species (Fig. 14.1). Choi et al. (2004a) reported a nearly perfect synteny between *M. truncatula* and alfalfa based on highly conserved nucleotide sequences. Although pea genome has one chromosome extra compared to *M. truncatula*, genomes of these two species has remarkably conserved co-linearity of genes and only interchromosomal

rearrangements have made differences between them (Choi et al. 2004b). Rearrangements in the chromosome 6 of *Medicago* differentiated *M. sativa* and *M. truncatula* (Choi et al. 2004b; Kalo et al. 2004). Interestingly, similar inter-chromosomal rearrangements were also identified between *M. truncatula* and chickpea. Though divergence between *M. truncatula* and *L. japonicus* was occurred >50 Mya, most of the genes in these two species have shown to be distributed over ~10 large blocks of synteny (Cannon et al. 2006).

Hougaard et al. (2008) reported similarity of intron-spanning markers of *Arabidopsis* with common bean and peanut. High level of macrosynteny (upto 85 %) was also established between cowpea and soybean on the basis of SNP discovered within EST data base of cowpea and found still high (82 %) between cowpea and *Medicago* (Muchero et al. 2009). Macrosynteny relationships based on >300 gene based Phaseolus loci derived from EST and BAC end sequences showed 55 synteny blocks on 35 of 40 chromosome arms of soybean used as reference sequences (McClean et al. 2010).

The macrosyntentic relationships were also studied on the basis of chromosome number. It was found that conservation of chromosome number (n = 11) across all of the phaseolid species have similar relative gene orders within corresponding chromosomes across all of species as observed between *Vigna radiata*, *Vigna unguiculata*, and *Phaseolus vulgaris* (Boutin et al. 1995; Choi et al. 2004a). Similarly in the hologalegina clade, all member of tribe Fabeae including *Vicia* (faba bean), *Pisum* (pea), *Lens* (lentil), *Lathyrus* (cicerchia) have n = 7 and *Cicer* (chickpea) of tribe Cicereae, has n = 8. The various clover species (typified by *M. truncatula*) are predominantly n = 8. Although *Medicago* and *Pisum* belong to separate tribes within the hologalegina, they share almost complete chromosome-scale synteny except *Medicago* chromosome 6 (Kalo et al. 2004) with limited synteny (Cannon et al. 2006).

Microsynteny

Microsyntenies have been first studied by making comparison between the *M. truncatula* and *Glycine max* genomes. A genome region surrounding at nematode-resistance gene (*rbg 1*) on chromosome 18 of *Gm* was studied (Mudge et al. 2005) and found that 75 % of genes are collinear between these two legume species. A hypersynteny region was also observed between these two species, which has 33 of 35 genes (94 %) conserved and collinear. One large synteny block between *Mt* 05 N (*M. truncatula*) and Lj 03-2S (*L. japonicus*) had 62 % genes which were collinear. Indeed synteny between *Mt* and *Lj* was found to extend nearly genome-wide, despite a time span of 40–50 Mya since their speciation (Cannon et al. 2006). More extensive microsynteny (up to 82 %) between *M. truncatula* and *L. japonicus* was also observed on the basis of 63 pairs of the sequenced BAC clones (Choi et al. 2004b). It has been shown that 54 % BAC contigs of soybean has some level of microsynteny with *M. truncatula* (Yan et al. 2003). Conserved gene order with at least six genes in common over 70 kb between *M. truncatula* and soybean has also been

identified around the genomic region of the putatively orthologous apyrase genes (Cannon et al. 2004). Within the syntenic interval, 14 (~48 %) distinct genes identified conserved in *M. truncatula* and soybean as well microsyntenic relationship between the chromosome CcLG06 of pigeonpea and chromose 1 of soybean (Varshney et al 2012). These studies on genome sequence alignment of different species and their comparison helped to understand the major macro-syntenic blocks and microsytenic genomic arrangements within blocks occurred during evolution of legume genome. Increasing the availability of sequences of more species will help to analyze actual changes occurred step by step in the genome of legume species.

Role of Whole Genome Duplication in Evolution of Legume Genome

Available genomic sequences of model plant species as well food legumes species make it possible to reconstruct the ancestral legume genome or at least the ancestral papilionoid genome. The genome sequences of the *Gm* (soybean), *Mt* (*Medicago*) and *Lj* (*Lotus*) indicated large scale architecture changes in the ancestral legume genome during the course of evolution of legume genome. A limited number of ancestral synteny blocks have been rearranged to generate present day papilinoid genomes. Comparison of all three genomes revealed 14 largely coherent blocks, which are nicely with apparent basal chromosome number of 7 for papilionoids (Polhill 1981). The syntenic relationship among legume species demonstrated a critical role of whole genome duplication (WGD) not only in evolution of plant species but also legume species (Doyle and Egan 2009). It was first studied in genome restructuring of soybean using restriction fragment length polymorphism. This study identified a homoeologous segment (paralogous sequences resulting from WGD) similar to nearly as long as whole chromosome (Shoemaker et al. 1996). Subsequently, available genomic sequences of different plant species revealed large scale duplication in ancient genome to generate present day plant genomes (Arabidopsis 2000; Jaillon et al. 2007; Tuskan et al. 2006; Blanc and Wolfe 2004; Pfeil et al. 2005; Shoemaker et al. 2006). These WGD events have been followed closely after the Cretaceous-Tertiary boundary event ~65 Mya and might probably be responsible for higher adaptability and greater tolerance to extreme conditions (Fawcett et al. 2009). Based on comparative studies, it was observed that each region of *Vitis vinifera* corresponds with three regions in other sequenced dicots. This suggested that an ancient event of triploidization occurred during 130–140 Mya was shared by many dicots (Jaillon et al. 2007).

During the recent past years, evolution of legume genome was studied in the light of such WGD events and subsequent further rearrangement in genome. Comparison of the genome sequences of *Mt* and *Lj* species revealed identification of dozen of duplicated synteny. After occurrence of major WGD event quickly significant genome rearrangements and gene loss were taken place before two species were diverged out. These significant changes in genome rapidly degraded the quality of

duplicate blocks. Based on synonymous substitution (Ks) estimates between paralongs (Blanc and Wolfe 2004; Pfeil et al. 2005; Schlueter et al. 2004) and topology of phylogebtic tree analysis (Cannon et al. 2006, 2010), it was identified that major WGD in the legume family was occurred ~58 Mya, which is before the separation of the *Mt* from Lj (~50 Mya) and splitting of these species (galegoid clade) from *Gm* (millettioid clade) that has occurred 54 Mya (Lavin et al. 2005). However there are different views of sharing this event for peanut and also for *Minossoudes* and *Caesealopodare* (Bertioli et al. 2009; Cannon et al. 2010). Use of ~126 cross-species EST mapped in *Arachis* and compared with available *Mt* and *Lj* sequences. Based on alignments of synteny blocks, they observed that papilionoid WGD event responsible for divergence of *Ararchis* from galegoids and phaseoloids occurred very early in the evolution of the subfamily (Bertioli et al 2009). Another study suggested that a WGD event occurred more recently (~13 Mya) has only responsible for splitting of soybean (Shoemaker et al. 1996) and further polyploidy in soybean has been estimated between 5 and 10 Mya. The WGD event causes two *Gm* blocks for each *Mt* genome and subsequently reshuffling have been taken place among duplicated *Glycine* genome segements due to simple as well as complex genome rearrangements (Young and Bharti 2012). Recently, comparison of pigeonpea genome sequences made on the basis of gene content and gene order with *Gm*, *Mt* and *Lj* suggested missing of some event in pigeonpea genome, which is member of same clade to which soybean belongs (Varshney et al. 2012). Overall, the whole-genome duplication occurred 58 Mya followed extensive rearrangements in the genome for stabilization before splitting the millettiod and galegoids clades i.e. 54 Mya. Some of these rearrangements have observed lineage specific, but based on microsyneny relationship between the pigeonpea and soybean, further local rearrangements were also occurred during the course of evolution (Varshney et al. 2012).

Impact of WGD Event on Evolution of Individual Genes and Gene Families

The WGD events obviously have a profound impact on genome architecture and an equally important role in the evolution of individual genes and gene families. These gene duplication events have been observed in plant genome as tandem, segmental and transposition (Freeling 2009). These duplicated genes were further categorized into various types on the basis of their function. When two duplicated genes are maintained and shared a function by spiting up the function of their ancestor is often called subfunctionalization (Force et al. 1999). While one of the duplicated genes takes part in a new function for various reasons is called neofunctionalization, although both genes are maintained (Lynch et al. 2001). Other possibility is that only one gene retained while its counterpart is deleted, is known as fractionation (Langham et al. 2004) or equivalently, diploidization. The duplication of each gene changes the entire evolutionary trajectory of a lineage into a noval direction and in legumes this has led a significant impact on nodulation and symbiosis with rhizobial bacteria (Young et al. 2011). Using the *Mt* genome sequence, relationship between

genome duplication and evolution of nodulation was established and a common ancestor has been presumed for nodulation in all the legumes belonging to a clade of rosids, Fabidae (Soltis et al. 1995). Based on multiple evidences, nodulation machinery predated the 58 Mya WGD and many of the known regulatory steps in rhizobial nodulation are shared with mycorrhizal signaling (Oldroyd and Downie 2008; Bonfante and Genre 2008). Only a few of the known recognition steps are exclusively associated with rhizobial nodulation. NEP, which is the key receptor-like kinage, is one of them (Oldroyd and Downie 2008) and have a homeolog, LYR1 in *Mt*. These duplicated genes were derived from WGD event occurred 58-Mya. NEP performs nodulation function in expression whereas LYR1 is responsible for the mycorrizal recognition functions in mycorrhizae (Gomez et al. 2009). In a nodulating nonlegume, *Parasponia andersonii*, a single gene has been identified to code a protein, which performs the functions of both NEP and LYr1 (Op den Camp et al. 2011). This supports that 58-Mya papilionoid WGD led to subfunctionalization of a more ancient gene carrying both functions. The example of sub-or neofunctionalization was observed that a separate nodulation-related transcription faction ERN1 also possesses a homeolog (ERN2) in *Mt*. These two homeologs have contrasting nodulation versus mycorrhizal expression patterns also derived from the 58 Mya WGD.

In legumes, 150 kb segment duplicated at two sites in *Mt* genome have maintained only 39 % genes which showed 81–85 % orthology of genes with *Gm*. Although fractionation is also taking place between the *Gm* paralogs, it has been taken place more slowly compared to *Mt* lineage as the result the numbers of gene pairs retained in *Gm* are much higher i.e. 69–100 % (Young and Bharti 2012). Impact of fractionation of duplicated genes was studied in evolution NBS-LRR disease-resistance gene family of legumes. Studying of a 1-Mb region centered around the resistance gene *Rpg1-b* located on *Gm*15 demonstrated that the homoeologous translocation of *Gm*15 region into the pericentromeric region of the chromosome has increased threefold physical size of *Gm*15. These duplicated regions although retained 77 % of duplicate genes, on the other hand, in these regions higher levels of fractionation of NBS-LRR genes was observed due to significant homoeologus-specific duplications and losses (Innes et al. 2008). In *Mt* genome, multiple linked genes (fractionation of duplicate genes in term of gene blocks) were retained in one duplicate but lost from other (Kim et al. 2009). Only differences in retrotransponson density between the two regions causes levels of structural variation and gene expressions. Difference in expression activity is significant because expression variation between retained gene pairs is an expectation of sub- and neo-functionalization.

Tandem duplication increases predicted genes in *L. japonicus* (12 %) and *M. truncatula* (17 %) and the majority of tandem duplication events are probably occurred independently after the divergence of the two species. Most of angiosperms (80 %) except legumes are likely to have a polyploid origin (Masterson 1994). Only soybean among legumes has long been known to be an ancient polyploid with putative homoeologous chromosomal regions (Shoemaker et al. 1996; Lee et al. 1999, 2001; Foster-Hartnett et al. 2002; Yan et al. 2003). Segmental duplications within the soybean genome were identified through fluorescence *in situ* hybridization of BACs (Pagel et al. 2004) and in the *M. truncatula* and *L. japonicus* genomes through high-throughput genome sequencing (Zhu et al. 2003).

Genomics in Analysis of Domestication Syndrome in Legumes

Genomics helped to study the syntenic relationship among the different species of legume taxa. During the evolution of legume genome, genomic regions responsible for traits modified from wild to cultivated species (usually known as domestication syndrome; Hammer 1984, 2003) are differed variably in term of rate and the order (Fuller 2007). For example, QTL for seed length and pod length on LG 7 in common beans are present in almost the same region on LG of azuki bean. However, QTL for pod and growth habit detected on LG7 in azuki bean were not detected on LG B5 of common bean. In another study, two and four QTL for seed weight were identified in the populations derived from wild × cultivated crosses in cowpea and mungbean, respectively (Fatokun et al. 1992). In this study, linkage groups established in these two crops shown a significant correspondence, and also showed correspondence for QTL detected for seed weight. Correspondence of QTLs for seed weight was also detected with azuki bean but with variable effects and some QTL for seed weight were observed crop specific. Though seed weight QTL appears to be conserved among these species over the linkage group, the main genome regions related to increased seed weight under domestication do not correspond among these related species. For example, in azuki bean, seed weight in cultivated taxa is about eight times that of the wild parent, while differences for seed weight in cultivated and wild parents of cowpea and mungbean has only been exhibited a fivefold (Fatokun et al. 1992). Azuki bean has the largest seed for the cultivated Asian *Vigna* (Tomooka et al. 2000). Therefore, different loci may be involved to increase seed size. Similarly, QTL for seed weight were also detected in tribes Phaseolae (i.e. soybean) and Vicieae (i.e. pea), which has corresponded with above crop species on the basis of genomic analysis (Maughan et al. 1996; Timmerman-Vaughan et al. 1996). These findings suggest that genomic region for seed weight has been conserved across the Leguminosae and play an important role in increasing seed size. Despite many parallels in the modifications during domestication between pea and common bean, none of genes involved in the domestication of both crops was identified so far. It has been observed that different genes are responsible for seed dispersal, growth habit, earliness, seed quality and seed pigmentation in pea and bean. However, many issues including seed dormancy, gigantism and the loss of photoperiod sensitivity that may involve homologous or orthologous sequences, which can be resolved by identifying the coding sequence of the gene affected in one crop followed by their mapping in the others. Thus genomics encourages to find out the ways of modification in characters occurred during the domestication of cultivated species.

References

Ahn S, Tanksley SD (1993) Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci U S A 90:7980–7984

Initiative AG (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Bertioli DJ, Moretzsohn MC, Madsen LH, Sandal N, Leal-Bertioli SC (2009) An analysis of synteny of *Arachis* with *Lotus* and *Medicago* sheds new light on the structure, stability and evolution of legume genomes. BMC Genomics 10:45

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell 16:1679–1691

Bonfante P, Genre A (2008) Plants and arbuscular mycorrhizal fungi: an evolutionary-developmental perspective. Trends Plant Sci 13:492–498

Boutin SR, Young ND, Olson TC, Yu ZH, Vallejos CE, Shoemaker RC (1995) Genome conservation among three legume genera detected with DNA markers. Genome 38:928–937

Cannon SB, Ilut D, Farmer AD, Maki SL, May GD (2010) Polyploidy did not predate the evolution of nodulation in all legumes. PLoS ONE 5:e11630

Cannon SB, May GD, Jackson SA (2009) Three sequenced legume genomes and many crop species: rich opportunities for translational genomics. Plant Physiol 151:970–977

Cannon SB, Mc Combie WR, Sato S, Tabata S, Denny R, Palmer L, Katari M, Young ND, Stacey G (2003) Evolution and microsynteny of the apyrase gene family in three legume genomes. Mol Genet Genomics 270:347–361

Cannon SB, Mitra A, Baumgarten A, Young ND, May G (2004) The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC Plant Biol 4:10

Cannon SB, Sterck L, Rombauts S, Sato S, Fl C (2006) Legume evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. Proc Natl Acad Sci U S A 103:14959–14964

Cannon SB, Sato S, Tabata S, Young ND, May GD (2011) Legumes as a model plant family. In: Pratap A, Kumar J (eds) Biology and breeding of food legumes. CAB International, Oxfordshire, pp 348–361

Choi HK, Kim D, Uhm T, Limpens E, Lim H, Mun JH, Kalo P, Penmetsa RV, Seres A, Kulikova O (2004a) A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. Genetics 166:1463–1502

Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR (2004b) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci U S A 101:15289–15294

Devos KM, Gale MD (2000) Genome relationships: the grass model in current research. Plant Cell 12:637–646

Doyle JJ, Luckow MA (2003) The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. Plant Physiol 131:900–910

Doyle JJ, Egan AN (2009) Dating the origins of polyploidy events. New Phytologist 186:73–85

Fatokun CA, Menancio-Hautea D, Danesh D, Young ND (1992) Evidence for orthologous seed weight genes in cowpea and mungbean based on RFLP mapping. Genetics 132:841–846

Fawcett JA, Maere S, Vandepeer Y (2009) Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. Proc Natl Acad Sci U S A 106:5737–5742

Force A, Lynch M, Pickett FB, Amores A, Yan YL (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545

Foster-Hartnett D, Mudge J, Larsen D, Danesh D, Yan H, Denny R, Penuela S, Young ND (2002) Comparative genomic analysis of sequences sampled from a small region on soybean (*Glycine max*) molecular linkage group G. Genome 45:634–645

Freeling M (2009) Bias in plant gene content following different sorts of duplication: tandem, whole genome, segmental, or by transposition. Annu Rev Plant Biol 60:433–453

Fuller DQ (2007) Contrasting patterns in crop domestication rates: recent archaeobotanical insights from the Old World. Ann Bot 100:903–924

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci U S A 95:1971–1974

Gepts P, Beavis WD, Brummer EC, Shoemakar RC, Stalker T, Weeden NF, Young ND (2005) Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through genomics conference. Plant Physiol 137:1228–1235

Gomez SK, Javot H, Deewatthanawong PD, Torres-Jerez I, Tang Y (2009) *Medicago truncatula* and *Glomus intraradices* gene expression in cortical cells harboring arbuscules in the arbuscular mycorrhizal symbiosis. BMC Plant Biol 9:10

Gualtieri G, Kulikova O, Limpens E, Kim DJ, Cook DR, Bisselin T, Geurts R (2002) Microsynteny between pea and Medicago truncatula in the SYM2 region. Plant Mol Biol 50:225–235

Hammer K (1984) The domestication syndrome. Kulturpflanze 3:11–34

Hammer K (2003) Evolution of cultivated plants and biodiversity. Nova Acta Leopoldina NF 87:133–146

Herendeen PS (1992) The fossil history of leguminosae from the eocene of southeastern North America. In: Herendeen PS, Dilcher DL (eds) Advances in legume systematics. Part 4. The fossil record. Kew, UK, Royal Botanic Gardens, pp 85–160

Herendeen PS and Wing S (2001) Papilionoid legume fruits and leaves from the Paleocene of northwestern Wyoming. Botany 2001 Abstracts, Botanical Society of America, USA (http://www.botany2001.org/). Accessed 4 Apr, 2013

Hougaard BK, Madsen LH, Sandal N, de Carvalho Moretzsohn M, Fredslund J (2008) Legume anchor markers link syntenic regions between *Phaseolus vulgaris*, *Lotus japonicus*, *Medicago truncatula* and *Arachis*. Genetics 179:2299–2312

Innes RW, Ameline-Torregrosa C, Ashfield T, Cannon E, Cannon SB (2008) Differential accumulation of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean. Plant Physiol 148:1740–1759

Jaillon O, Aury JM, Nöel B, Policriti A, Clepet C (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467

Kalo P, Seres A, Taylor SA, Jakab J, Kevei Z (2004) Comparative mapping between *Medicago sativa* and *Pisum sativum*. Mol Genet Genomics 272:235–246

Kim KD, Shin JH, Van K, Kim DH, Lee SH (2009) Dynamic rearrangements determine genome organization and useful traits in soybean. Plant Physiol 151:1066–1076

Kumar J, Pratap A, Solanki RK, Gupta DS, Goyal A, Chaturvedi SK, Kumar S (2011) Advances in genomics resources for improving food legume crops. J Agric Sci 150:289–318

Langham RJ, Walsh J, Dunn M, Ko C, Goff SA (2004) Genomic duplication, fractionation and the origin of regulatory novelty. Genetics 166:935–945

Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. Syst Biol 54:575–594

Lee JM, Bush A, Specht JE, Shoemaker RC (1999) Mapping duplicate genes in soybean. Genome 42:829–836

Lee JM, Grant D, Vallejos CE, Shoemaker RC (2001) Genome organization in dicots. II. Arabidopsis as a bridging species to resolve genome duplication events among legumes. Theor Appl Genet 103:765–773

Lewis G, Schrire B, Mackind B, Lock M (2005) Legumes of the world. Royal Botanic Gardens, Kew

Lynch M, O'Hely M, Walsh B, Force A (2001) The probability of preservation of a newly arisen gene duplicate. Genetics 159:1789–1804

Masterson J (1994) Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science 264:421–424

Maughan PJ, Saghai MMA, Buss GR (1996) Molecular-marker analysis of seed-weight: genomic locations, gene action, and evidence for orthologous evolution among three legume species. Theor Appl Genet 93:574–579

McClean PE, Mamidi S, McConnell M, Chikara S, Lee R (2010) Synteny mapping between common bean and soybean reveals extensive blocks of shared loci. BMC Genomics 11:184

Menancio-Hautea D, Fatokun CA, Kumar L, Danesh D, Young ND (1997) Comparative genome analysis of mungbean (*Vigna radiata* (L.) Wilczek) and cowpea (*V. unguiculata* (L.) Walpers) using RFLP mapping data. Theor Appl Genet 86:797–810

Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S (2009) A consensus genetic map of cowpea [*Vigna unguiculata* (L) Walp.] and synteny based on EST-derived SNPs. Proc Natl Acad Sci U S A 106:18159–18164

Mudge J, Cannon SB, Kalo P, Oldroyd GE, Roe BA, Town CD, Young ND (2005) Highly syntenic regions in the genomes of soybean, *Medicago truncatula*, and *Arabidopsis thaliana*. BMC Plant Biol 5:15

Nayak SN, Zhu H, Varghese N, Datta S, Hk C (2010) Integration of novel SSR and gene-based SNP marker loci in the chickpea genetic map and establishment of new anchor points with *Medicago truncatula* genome. Theor Appl Genet 120:1415–1441

Oldroyd GE, Downie JA (2008) Coordinating nodule morphogenesis with rhizobial infection in legumes. Annu Rev Plant Biol 59:519–546

Op den Camp RHM, Streng A, De MS, Cao Q, Polone E (2011) LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume Parasponia. Science 331:909–912

Pagel J, Walling JG, Young ND, Shoemaker RC, Jackson SA (2004) Segmental duplications within the *Glycine max* genome revealed by fluorescence in situ hybridization of bacterial artificial chromosomes. Genome 47:764–768

Pan AD, Jacobs BF, Herendeen PS (2010) Detarieae sensu lato. (Fabaceae) from the Late Oligocene (27.23 Ma) Guang River flora of north-western Ethiopia. Bot J Linnean Soc 163:44–54

Pfeil BE, Schlueter JA, Shoemaker RC, Doyle JJ (2005) Placing paleopolyploidy in relation to taxon divergence: a phylogenetic analysis in legumes using 39 gene families. Syst Biol 54:441–454

Polhill RM (1981) Papilionoideae. In: Polhill RM, Raven PH, (eds) Advances in legume systematics. Part 1. Royal Botanic Gardens, Kew, pp 191–204

Polhill RM, Raven PH (1981) Advances in legume systematics. Part 1. Royal Botanic Gardens, Kew

Raven PH, Axelrod DI (1974) Angiosperm biogeography and past continental movements. Annal Miss Bot Garden 61:539–657

Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T (2008) Genome structure of the legume, *Lotus japonicus*. DNA Res 15:1–8

Schlueter JA, Dixon P, Granger C, Grant D, Clark L (2004) Mining EST databases to resolve evolutionary events in major crop species. Genome 47:868–876

Schrire BD, Lavin M, Lewis GP (2005a) Global distribution patterns of the Leguminosae: insights from recent phylogenies. In: Friis I, Balslev H (eds) Plant diversity and complexity patterns: local, regional and global dimensions, vol 55, Biologiske Skrifter. Special-Trykkeriet Viborg A/S, Viborg, Denmark, pp 375–422

Schrire BD, Lewis GP, Lavin M (2005b) Biogeography of the leguminosae. In: Lewis G, Schrire G, Mackinder B (eds) Legumes of the world. Royal Botanic Gardens, Kew, pp 21–54

Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC (1996) Genome duplication in soybean (Glycine subgenus soja). Genetics 144:329–338

Shoemaker RC, Schlueter J, Doyle JJ (2006) Paleopolyploidy and gene duplication in soybean and other legumes. Curr Opin Plant Biol 9:104–109

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Simon CJ, Muehlbauer FJ (1997) Construction of a chickpea linkage map and its comparison with maps of pea and lentil. J Hered 88:115–119

Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC (1995) Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proc Natl Acad Sci U S A 92:2647–2651

Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF, Axtell M, Swenson SM, Prince LM, Kress WJ, Nixon KC, Farris JS (2000) Angiosperm phylogeny inferred from 18S rDNA, rbcL, and atpB sequences. Bot J Linnean Soc 133:381–461

Timmerman-Vaughan GM, McCallum JA, Frew TJ, Weeden NF, Russell AC (1996) Linkage mapping of quantitative trait loci controlling seed weight in pea (*Pisum sativum* L.). Theor Appl Genet 93:431–439

Tomooka N, Kashiwaba K, Vaughan DA, Ishimoto M, Egawa Y (2000) The effectiveness of evaluating wild species: searching for sources of resistance to bruchid beetle in the genus *Vigna* subgenus Ceratotropis. Euphytica 115:27–41

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596–1604

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK (2012) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotechnol 30:83–89

Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'an B, Millan T, Zhang X, Ramsay LD, Iwata A, Wang Y, Nelson W, Farmer AD, Gaur PM, Soderlund C, Penmetsa RV, Xu C, Bharti AK, He W, Winter P, Zhao S, Hane JK, Carrasquilla-Garcia N, Condie JA, Upadhyaya HD, Lu MC, Thudi M, Gowda CLL, Singh NP, Lichtenzveig J, Gali KK, Rubio J, Nadarajan N, Dolezel J, Bansal KC, Xu X, Edwards D, Zhang G, Kahl G, Gil J, Singh KB, Datta SK, Jackson S, Wang J, Cook DR (2013) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. Nat Biotechnol 31:240–246

Wing SL, Herrera F, and Jaramillo C (2004) A Paleocene flora from the Cerrajón Formation, Guajíra Peninsula, northeastern Colombia. In: VII International organization of paleobotany conference abstracts, 21–26 March, Museo Egidio Feruglio, Trelew, Argentina, pp 146–147

Yan HH, Mudge J, Kim D-J, Shoemaker RC, Cook DR, Young ND (2003) Estimates of conserved microsynteny among the genomes of *Glycine max*, *Medicago truncatula*, and *Arabidopsis thaliana*. Theor Appl Genet 106:1256–1265

Yan HH, Mudge J, Kim DJ, Shoemaker RC, Cook DR, Young ND (2004) Comparative physical mapping reveals features of microsynteny between *Glycine max*, *Medicago truncatula*, and *Arabidopsis thaliana*. Genome 47:141–155

Young N, Debellé F, Oldroyd G, Geurts R, Cannon SB (2011) The *Medicago* genome provides insight into the evolution of rhizobial symbioses. Nature 480:520–524

Young ND, Bharti AK (2012) Genome-enabled insights into legume biology. Annu Rev Plant Biol 63:283–305

Young ND, Mudge J, Ellis TH (2003) Legume genomes, more than peas in a pod. Curr Opin Plant Biol 6:199–204

Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C, Zhang J, Zhang Y, Li R, Xu Z, Li S, Li X, Zheng H, Cong L, Lin L, Yin J, Geng J, Li G, Shi J, Liu J, Lv H, Li J, Wang J, Deng Y, Ran L, Shi X, Wang X, Wu Q, Li C, Ren X, Wang J, Wang X, Li D, Liu D, Zhang X, Ji Z, Zhao W, Sun Y, Zhang Z, Bao J, Han Y, Dong L, Ji J, Chen P, Wu S, Liu J, Xiao Y, Bu D, Tan J, Yang L, Ye C, Zhang J, Xu J, Zhou Y, Yu Y, Zhang B, Zhuang S, Wei H, Liu B, Lei M, Yu H, Li Y, Xu H, Wei S, He X, Fang L, Zhang Z, Zhang Y, Huang X, Su Z, Tong W, Li J, Tong Z, Li S, Ye J, Wang L, Fang L, Lei T, Chen C, Chen H, Xu Z, Li H, Huang H, Zhang F, Xu H, Li N, Zhao C, Li S, Dong L, Huang Y, Li L, Xi Y, Qi Q, Li W, Zhang B, Hu W, Zhang Y, Tian X, Jiao Y, Liang X, Jin J, Gao L, Zheng W, Hao B, Liu S, Wang W, Yuan L, Cao M, McDermott J, Samudrala R, Wang J, Wong GK, Yang H (2005) The genomes of *Oryza sativa*: a history of duplications. PLoS Biol 3:e38

Zhu H, Choi HK, Cook DR, Shoemaker RC (2005) Bridging model and crop legumes through comparative genomics. Plant Physiol 137:1189–1196

Zhu H, Kim DJ, Baek JM, Chi HK, Ellis LC, Kuester H, McCombie WR, Peng HM, Cook DR (2003) Syntenic relationships between *Medicago truncatula* and *Arabidopsis* reveal extensive divergence of genome organization. Plant Physiol 131:1018–1026

Chapter 15
Advances in Pea Genomics

Petr Smýkal and Eva Konečná

Abstract Pea (*Pisum sativum* L.) is important temperate region pulse, with feed, fodder and vegetable uses. Pea was the model organism for Mendel´s discovery of the laws of inheritance, making it the foundation of modern genetics. However, subsequent progress in pea genomics has lagged behind many other plant species, largely as a consequence of its genome size (4.45 Gbp), consisting to large part (75–97 %) of repetitive sequences and its relatively low economic significance (compare to cereals or soybean). There is a long history of genetic mapping studies in pea leading to seven linkage groups definition as well as subsequent marker/trait association. The availability of the genome sequences of three phylogenetically related legume species (*Medicago truncatula*, *Lotus japonicus* and *Glycine max*) has offered opportunities for genome wide comparison. Combination of a candidate gene and colinearity approach has allowed the identification of genes underlying agronomicaly important traits. The progress in the understanding of genes and functional association to traits in the model legume species has been accompanied by gene-based marker development in pea. This together with enhanced computational power and access to diverse germplasm collections, led to association mapping application to identify genetic variation related to desirable agronomic traits. Some of this knowledge has already been applied to marker assisted selection (MAS) programs, increasing the precision and shortening the breeding cycle. Comprehensive pea genomic resources already exist and include several types of molecular marker sets as well as both transcriptome and proteome datasets. Various available marker types have been used for assessment of genetic diversity and mapping of agronomicaly important Quantitative Trait Loci (QTL). Fast neutron and TILLING pea

P. Smýkal, Ph.D. (✉)
Department of Botany, Palacky University in Olomouc, Slechtitelu 11, Olomouc 783 71, Czech Republic
e-mail: petr.smykal@upol.cz

E. Konečná, Ing.
Department of Plant Science, Mendel University in Brno, Brno, Czech Republic

S. Gupta et al. (eds.), *Legumes in the Omic Era*, DOI 10.1007/978-1-4614-8370-0_15,

mutant populations are available for reverse genetics approaches, BAC libraries for positional gene cloning as well as transgenic and *in vitro* regeneration for proof of function or novel gene incorporation. Current genomic knowledge and technologies can facilitate the allele mining for novel traits and incorporation from wild Pisum sp. into elite domestic backgrounds. The availability of high throughput sequencing, genotyping and the -omics methodologies, these hold great promise for the development of novel, highly accurate selective breeding tools for improved pea genotypes suitable for various climates and farming systems.

Keywords Pea • *Pisum sativum* • Genomics • Marker-assisted breeding • QTLs

Introduction

Dry pea ranks second after *Phaseolus* as the most widely grown grain legume in the world with global production of 10.4 M tonnes in 2009 (FAOSTAT 2012). It is important temperate region pulse, with feed, fodder and vegetable uses. Its seeds are rich in protein (23–25 %), slowly digestible starch (50 %), soluble sugars (5 %), fibre, minerals and vitamins as well as in secondary metabolites such as isoflavonoids with anticancer and other health-promoting activities (Bastianelli et al. 1998; Dixon and Sumner 2003).

Pea is one of the world's oldest domesticated crops. Archaeological evidence dates the existence of pea back to 8000 BC in Near East and in Europe where it has been found since the Stone and Bronze Ages (De Candolle 1882; Vavilov 1949; Smartt 2005; Ambrose 1995; Zohary and Hopf 2000; Abbo et al. 2010; Upadhyaya et al. 2011). The center of pea genetic diversity is the broad area of Fertile Crescent through Turkey, Syria, Iraq, Israel and Lebanon. It extends further east to Central Asia (Iran, Afghanistan, Pakistan and Turkmenistan) (Smýkal et al. 2011). Ethiopia is considered as a secondary center of diversity (Van der Maesen 1998).

Pea belongs to genus *Pisum* which comprises of mainly three species i.e. *P. sativum* L. with subsp. *sativum* (includes var. *sativum* and var. *arvense*), subsp. *elatius*, *P. fulvum* and *P. abyssinicum*. The most used classification is of (Maxted and Ambrose 2000) to which to which *Vavilovia formosa* was added to classify four species (Smýkal et al. 2011). Primary gene pool consists of *P. sativum* including wild *P. sativum* subp. *elatius*, a secondary gene pool is composed of *P. fulvum* and a tertiary gene pool consisting only of *Vavilovia formosa*.

Genome Size

Nuclear genome size estimates have been produced for several accessions of pea using different methods and estimated to be 9.09 pg DNA/2C corresponding to the haploid genome size (1C) of 4.45 Gbp (Dolezel and Greilhuber 2010). The average GC content is 37.4 % and approx. 30 % C residues are methylated (Pradhan and

Adams 1995). Early studies of sequence composition of the pea genome employing DNA reassociation kinetics and melting behavior measurements indicated that its large part (75–97 %) is made up of repetitive sequences, being confirmed recently by next generation sequencing approach (Macas et al. 2007). TheTy3/gypsy LTR-retrotransposons were identified as the main component of the pea repeats, with highly amplified group of Ogre elements alone representing 20–33 % of the pea genome. Another interesting lineage of copia-type Angela-family retrotransposon has been shown to be evolutionary conserved and involved in microsatellite repeat dispersal (Smýkal et al. 2009). Some of these elements were found useful as a source of molecular or cytogenetic markers allowing discrimination of individual chromosomes within the karyotype or diversity studies (Flavell et al. 2003; Jing et al. 2005; Neumann et al. 2001; Smýkal 2006).

Genomic Resources

Genetic Maps and Mapping Populations

There is a long history of genetic mapping studies in pea (McPhee 2007; Aubert et al. 2006). Different types of markers were successively used: morphological, isozymes, RFLP, RAPD, SSR, ESTs and finally gene-based (Blixt 1972; Hall et al. 1997; Aubert et al. 2006; Weeden and Marx 1987; Irzykowska et al. 2001; Laucou et al. 1998; Loridon et al. 2005; Gilpin et al. 1997; Weeden and Boone 1999; Timmerman-Vaughan et al. 2000; Konovalov et al. 2005; Deulvot et al. 2010; Bordat et al. 2011).

Combining these markers, consensus map was built from a population of 51 RIL derived from JI1794 × Slow (Weeden et al. 1998) (Table 15.1). Later, pea consensus linkage maps were obtained using three connected RIL populations (JI15 × JI1194, JI15 × JI399 and JI281 × JI399) (Hall et al. 1997) (Table 15.1). The total length of the integrated map (937 cM) was close to the expectation from chiasma distribution (Hall et al. 1997). Three different crosses (Terese × K586-RIL1, Champagne × Terese-RIL2, Shawnee × Bohatyr) were used to build a composite genetic map of 1,430 cM comprising 239 microsatellite markers, till now the most cited and used pea genetic map (Loridon et al. 2005) (Table 15.1). The markers are quite evenly distributed throughout the seven linkage groups of the map, with 85 % of intervals between the adjacent SSR markers being smaller than 10 cM. This map was used to localize numerous QTLs for disease resistance as well as quality and morphological traits. Later, maps composed of genes of known function, were developed, such as consensus genetic map from the populations RIL1 and RIL2 which covered 1,458 cM and comprised 363 loci (Aubert et al. 2006) (Table 15.1). The last consensus map published in pea provides the most comprehensive view of the pea map until now (Bordat et al. 2011) (Table 15.1). This map was built from data obtained for 1,022 RIL belonging to four RIL populations (VavD265 × Cameor, Ballet × Cameor, VavD265 × Ballet, China × Cameor). It includes 214 functional markers, representing genes from diverse functional classes such as development, carbohydrate

Table 15.1 Molecular maps developed in pea

Population	Number of lines	Type and number of marker used	Mean spacing (cM)	Genetic length (cM)	Reference
Multiple	Not given	Establishment of 7 linkage groups, 37 morphological	Not given	Not given	Lamprecht (1948)
Multiple	Not given	Assiggment of linkage groups to chromosomes	Not given	Not given	Lamprecht (1961)
Multiple	Not given	128 loci	Not given	Not given	Lamprecht (1974)
Multiple	Not given	128 loci of lamprecht and 44 new genes	Not given	Not given	Blixt (1977)
Multiple	Not given	27 isozyme loci	Not given	Not given	Weeden (1985)
JI1794×Slow	51	Concensus linkage map, 465 RFLP, isozymes, RAPD, SSR, STSM	2	800	Weeden (1998)
JI281×JI399	71	151 RFLP, morphological and rDNA and storage protein genes	Not given	1,700	Ellis et al. (1992)
JI15×JI61	38	51 RFLP	Not given	1,700	Ellis et al. (1992)
JI15×JI1194	50	65 RFLP	Not given	1,700	Ellis et al. (1992)
JI813×JI1201	48	13 RFLP	Not given	1,700	Ellis et al. (1992)
JI1794×Slow	51	Classical mutants, isozymes, RFLP and EST	Not given	1,700	Gilpin et al. (1997)
Primo×OSU 442-15	102 F_2	209 EST, RFLP, RAPD, AFLP	Not given	1,330	Gilpin et al. (1997)
JI15×JI1194	F_1, F_2	120 Markers	Not given	937	Hall et al. (1997a, b)
JI15×JI399	F_1, F_3	121 Markers	Not given	938	Hall et al. (1997a, b)
JI281×JI399	F_1, F_4	122 Markers	Not given	939	Hall et al. (1997a, b)
Terese×K586	139	247 RAPD markers on nine LG, 9 RFLP, 2 morphological (le, af),	Less than 30	1,139	Laucou et al. (1998)
JI281×JI399	71	355 RAPD markers	Less than 30	1,881	Laucou et al. (1998)
Primo×OSU 442-15	102	209 EST, RFLP, RAPD, AFLP	Not given	1,330	Timmerman-Vaughan et al. (2000)
JI1794×Slow	51	235 RAPD, RFLP, AFLP	Not given	1,289	Timmerman-Vaughan et al. (2000)
Rovar×3176-A26	148	84 RFLP, RAPD	Not given	Not given	Timmerman-Vaughan et al. (2000)

JI1794 × Slow, MN313 × JI1794	53, 51	37 gene-based, orthologous to *Lens*	Not given	Not given	Brauner et al. (2002)
Wt10245 × Wt11238	116	140 AFLP, RAPDs, ISSRs, CAPSs or STSs plus 30 anchor morphological, isozyme and STS markers	10	2,416	Irzykowska and Wolko (2004)
Medicago sativa F_2 and pea RILs from Ellis et al. (1992), Hall et al. (1997)		Genetic mapping of 74 gene-specific	Not given	Not given	Kalo et al. (2004)
Wt11238 × Wt3557	116 F_2/ 108 F_4	AFLP, RAPD, ISSR, CAPS, STS, 10 morphological, 10 isozymes	22	2,172	Gawlowska et al. (2005)
Wt10245 × Wt11238	116	AFLP, RAPD, ISSR, CAPS, STS, 10 morphological, 10 isozymes	10	841	Gawlowska et al. (2005)
Terese × K586	139 RIL	216 anonymous and 13 genic SSR	6	1,430	Loridon et al. (2005)
Champagne × Terese	164 RIL	216 anonymous and 13 genic SSR	6	1,430	Loridon et al. (2006)
Shawnee × Bohatyr	187 RIL	216 anonymous and 13 genic SSR	6	1,430	Loridon et al. (2007)
Multiple	–	242 gene-based	Not given	Not given	Weeden (2005)
Terese × K586	139	363 loci, of that 7 SSR, 77 gene-based syntenical to *Medicago truncatula*, 63 SNPs, 5 morphological	4	1,458	Aubert et al. (2006)
Champagne × Terese	164	363 loci, of that 7 SSR, 77 gene-based syntenical to *Medicago truncatula*, 63 SNPs, 5 morphological	4	1,458	Aubert et al. (2006)
China × Cameor	90	384 SNP markers represent 205 different genes, mapped 110 of them	8	680	Deulvot et al. (2010)
Terese × K586	139	180 SSR, 133 RAPD, 214 gene based, 48 known mutations and 15 protein, 536 markers and 5,460 unigenes	97 % of marker intervals below 10	1,389	Bordat et al. (2011)

(continued)

Table 15.1 (continued)

Population	Number of lines	Type and number of marker used	Mean spacing (cM)	Genetic length (cM)	Reference
Champagne × Terese	164	180 SSR, 133 RAPD, 214 gene based, 48 known mutations and 15 protein, 536 markers and 5,460 unigenes in silico	97 % of marker intervals below 10	1,389	Bordat et al. (2011)
VavD265 × Cameor	211	180 SSR, 133 RAPD, 214 gene based, 48 known mutations and 15 protein, 536 markers and 5,460 unigenes in silico	97 % of marker intervals below 10	1,389	Bordat et al. (2011)
Ballet × Cameor	207	180 SSR, 133 RAPD, 214 gene based, 48 known mutations and 15 protein, 536 markers and 5,460 unigenes in silico	97 % of marker intervals below 10	1,389	Bordat et al. (2011)
VavD265 × Ballet	211	180 SSR, 133 RAPD, 214 gene based, 48 known mutations and 15 protein, 536 markers and 5,460 unigenes in silico	97 % of marker intervals below 10	1,389	Bordat et al. (2011)
China × Cameor	90	180 SSR, 133 RAPD, 214 gene based, 48 known mutations and 15 protein, 536 markers and 5,460 unigenes in silico	97 % of marker intervals below 10	1,389	Bordat et al. (2011)
Early Freezer 680 × PI 269818	68	45 new EST-SSR	Not given	Not given	Decarie et al. (2012)
C580257A × 74SN3A	87	45 new EST-SSR	Not given	Not given	Decarie et al. (2012)

metabolism, amino acid metabolism, transport and transcriptional regulation. It also includes 180 SSR, 133 RAPD and 3 morphological markers, thus is intrinsically related to previous maps (Hall et al. 1997; Aubert et al. 2006; Laucou et al. 1998; Loridon et al. 2005; Gilpin et al. 1997; Weeden et al. 1998). Based on markers shared with previously published maps, 48 known mutations and 15 protein or gene markers could be placed onto this consensus map. Moreover, this map provides basis for translational genomic approaches among legumes. Comparing these different maps, length variations or gene order discrepancies among populations were observed. One contribution to excess map length could be miscoring due to DNA methylation, or deficiency of heterozygotes (Knox and Ellis 2001, 2002). Different recombination rates could also occur, with longer maps for crosses among close genotypes and shorter maps for wider crosses. Gene order discrepancies could be also due to translocations in the different RIL population used or/and by missing data for markers genotyped in one population (Bordat et al. 2011; Ellis and Poyser 2002).

DNA Based Molecular Markers

Microsatellite Markers

Microsatellite markers, also termed simple sequence repeats (SSR), occur ubiquitously in eukaryotic genomes and were among the first markers used for mapping studies. A common set of Simple Sequence Repeats (SSRs or microsatellites) was developed (Loridon et al. 2005; Burstin et al. 2001; Ford et al. 2002) and broadly used for mapping. SSR have also been popular for assessing *Pisum* diversity because of their high polymorphism and information content, co-dominance and reproducibility (Smýkal et al. 2008a; Zong et al. 2009; Ford et al. 2002; Baranger et al. 2004; Tar'an et al. 2005). On the other hand, microsatellites have much higher mutation rates (estimated at 10–4 site^{-1} year^{-1}) than the nucleotide substitution rate and therefore suffer from homoplasy (the state when identical alleles have arisen by two or more different pathways of descent) in very diverse material (Cieslarová et al. 2011). The high diversity in the Pisum genus suggests that the risk of homoplasy in wide surveys of pea germplasm using microsatellites might be high (Smýkal et al. 2012a; Ellis 2011). Most recently, expressed sequence tag (EST)-derived simple sequence repeat (eSSR) markers have become an important resource for gene discovery and comparative mapping studies (Cieslarová et al. 2011; Mishra et al. 2012; Decarie et al. 2012).

Retrotransposone Based Markers

Other marker types broadly used for diversity studies include retrotransposon-based methods, such as Sequence Specific Amplified Polymorphism (SSAP), Inter-Retrotransposon Amplified Polymorphism (IRAP) and Retrotransposon-based Insertion Polymorphism (RBIP) (Flavell et al. 2003; Smýkal 2006; Ellis et al. 1998; Vershinin et al. 2003). The later method was developed into high-throughput

genotyping technology based on insertion/deletion of retrotransposons from the Ty1-copia retrotransposable element sequences and used extensively in studies of molecular variation among pea core collections (Smýkal et al. 2008a, b, 2011; Jing et al. 2010, 2012). It was shown to provide greater power for phylogeny and genetic relationship studies and is therefore suited for in-depth phylogeny and germplasm diversity studies (Smýkal et al. 2008b, 2011; Jing et al. 2010, 2012). Another Angela-family class of highly abundant retrotransposons in pea was identified and used for fingerprinting (Smýkal et al. 2009). The list of pea markers published up to now allow for QTL or gene mapping, marker assisted selection (MAS), genetic diversity survey, and association studies. Although SSR and RBIP marker types are still widespread, they will probably be replaced by SNP or sequencing genotyping. Several pea cDNA sequencing projects are ongoing and should provide, at last, a complete Unigene set for the pea RNA-Seq gene atlas, INRA, France, Cool Season Food Legume Genome Database, USDA, USA, http://www.gabcsfl.org (Franssen et al. 2011). This set should facilitate the investigation of whole genome sequence polymorphisms in pea and pave the way to Genome Wide Association Studies (GWAS) as well as Genomic Selection (GS) providing link between phenotype and genotype. Improvements in marker methods have been accompanied by refinements in computational methods to convert original data into useful representation of diversity and genetic structure. Initially and still used distance-based methods have been challenged by model-based Bayesian approaches. The incorporation of probability, measures of support, ability to accommodate complex model and various data types make them more attractive and powerful (Reif et al. 2005; Beaumont and Rannala 2004; Corander et al. 2003).

Functional: cDNA/EST Markers

The expressed sequence tag (EST) approach identifies candidate genes via cDNAs (copies of mRNA) transcribed in response to a particular stimuli/trait of interest. The availability of rapidly growing sequence databases allows the detection of regions showing sequence similarities in functionally related gene products from distantly related organisms. Thus it is increasingly possible to assign putative functions for a large proportion of anonymous cDNA clones/ESTs. Consequently EST analysis is a rapid and efficient way for providing preliminary information of the expressed profiles for the most abundant transcripts of genes in any particular tissue in different physiological conditions, and thus the identification of regulatory genes. Moreover in case of large and complex genomes, such of pea, ESTs allow fast dissection of gene coding regions and elimination of repetitive sequences. The initial set of pea ESTs was developed by Gilpin et al. 1997. Large online pea database "CROP-EST" (http://pgrc.ipk-gatersleben.de/cr-est/) included two pea cDNA libraries with a total of 9,377 ESTs (Künne et al. 2005). Recently, a large database comprising 7,610 unique genes from shoot apical meristem cDNA libraries was generated (Liang et al. 2009) and even larger RNA-seq atlas (http://bios.dijon.inra.fr/FATAL/cgi/pscam.cgi).

Single Nucleotide Polymorphism (SNP) Markers

Single Nucleotide Polymorphism (SNP) markers are currently preferentially used to assess genetic diversity, for genetic mapping and for tagging alleles of functional interest. Gene-specific studies have identified SNP markers useful for characterising diversity and relationships among pea genotypes (Jing et al. 2007). When SNP markers are incorporated into arrays, they offer assessment of thousands of gene-related SNPsSNP in a single reaction. Thirty seven genes were mapped and used in 384-SNP array for pea from a diverse genotypic background (Deulvot et al. 2010). In comparison to other types of markers, the rate of SNP discovery is almost unlimited. In pea, one SNP was detected per 94 bp on average (i.e. one in 165 bp in coding regions and one in 60 bp in non-coding regions) (Aubert et al. 2006; Jing et al. 2007). Recently, a set of SNP markers using Illumina Veracode genotyping technology has been in pea and used to build consensus map (Deulvot et al. 2010; Bordat et al. 2011). Within Crop EST project, 9,377 ESTs with BLAST identified 8,238 ESTs were obtained from two cDNA libraries of pea (http://pgrc.ipk-gatersleben.de/cr-est) (Künne et al. 2005). Recently, a set of 37,455 contig sequences were assembled from 3,084,253 high quality 454 reads (1.2 Gbp) of pea variety Aragorn using the Newbler algorithm. These 37,368 putative transcripts were of average length 1,045 bp and represent 25,353 isotigs, and represent 34,846 unigenes, 8,817 contigs and 26,029 singletons (http://www.gabcsfl.org). Similarly, INRA Dijon, France has produced pea RNA-Seq Gene Atlas (http://bios.dijon.inra.fr/FATAL/cgi/pscam.cgi). This web-portal provides the first full-length Unigene set expression atlas for pea. Twenty pea cDNA libraries were prepared from different above- and below- ground cv. "Cameor" plant organs, at different stages, and for different nutrition conditions. Libraries were sequenced using Next-Generation Sequencing technologies. Sequences were assembled de novo and a full-length Unigene set was produced. The sequencing depth of each cDNA contig relates to the expression level of transcripts. This gene atlas presents the pattern of expression and thus provides useful functional information for each cDNA contig (Alves-Carvalho et al. in prep). In the future, new RNA-Seq experiments will be added to this portal to enlarge the scope of the atlas.

Comparative and Functional Genomics

Comparative genetic analysis among legumes was first presented by Vavilov's studies on homologues series of similar heritable variation in related Vicieae tribe species (Vavilov 1922). The first molecular evidence of macrosynteny between legumes was given by the comparison of genetic maps of economically important legumes: between pea and lentil, pea and chickpea, pea and *Medicago*, as well as among several legume species (Aubert et al. 2006; Weeden et al. 1992; Simon and Muehlbauer 1997; Choi et al. 2004; Kalo et al. 2004) Recently, cross species gene-based markers were used to identify homologous genome segments among eight

legume species (*M. truncatula*, alfalfa, *L. japonicus*, chickpea, soybean, mungbean, common bean and pea) with *M. truncatula* used as reference genome for consensus map (Cannon et al. 2009; Choi et al. 2004). By systematically searching for the best homologues of the genes mapped in the pea consensus map in sequenced legume genomes, comparison of gene orders in pea, *M. truncatula*, *L. japonicus* and soybean was made and specified the overall conservation of gene order and the correspondence among pea linkage groups and *M. truncatula* and *L. japonicus* pseudo chromosomes (Bordat et al. 2011). More data will help refine observed synteny. Comparative mapping also allows for investigating the paleo-history of the pea genome. Scenario of evolution of the seven pea chromosomes from the paleo-hexaploid ancestor of Eudicot was proposed (Bordat et al. 2011). Finally, translational genomics is beginning to assist in identifying candidate genes or saturating markers in a zone of interest of pea (Wang et al. 2008; Hecht et al. 2007, 2011; Hellens et al. 2010).

Pea Transcriptome and Proteome Analysis

Transcriptome analysis has been a key area of biological investigation for decades. In the absence of a completely sequenced genome, EST collections, such as unigenes at NCBI (NCBI 2013), or tentative consensus sequences at DFCI (DFCI Plant Gene Indices. http://compbio.dfci.harvard.edu/tgi/plant.html) produced by traditional Sanger sequencing have proven extremely useful for research. The development of expressed sequence tags (ESTs) from pea has provided a source for mining novel simple sequence repeats (SSR) markers, valuable resources for gene discovery, expression analysis, and genome annotation (Kaur et al. 2012; Gong et al. 2010). Pea 6 k oligo-array (Ps6kOLI1) developed from diverse sources of genomic sequence, especially seed EST libraries, have been performed for several transcriptome analyses. Seed development processes and specific genes involved in primary metabolism or hormone deficiency were investigated (Weigelt et al. 2008, 2009; Riebeseel et al. 2010; Radchuk et al. 2010). Hydrogen peroxide has been shown to accumulate during seed germination. The effect of treatment of mature pea seeds with hydrogen peroxide on several oxidative features and the expression of genes known to be activated by hydrogen peroxide were monitored as well as metabolites and function of antioxidant enzymes during maturations of pea fruits (Barba-Espín et al. 2011; Matamoros et al. 2009). The development of transcription quantitative PCR methods facilitated transcript detection, increased the experimental throughput, and reduced the required quantity of input RNA. Expression and quantitative studies were used to explore the interactions between the hormones and unifololiata (UNI) gene that control leaf morphogenesis. It was shown that rate of increase in leaf complexity during shoot ontogeny and adult leaf complexities are controlled by gibberellic acid through UNI gene (DeMason and Chetty 2011). Important evaluation of candidate reference genes in pea varieties subjected to various abiotic and biotic stresses was undertaken, resulting in identification of tubulin-3 and TIF genes as the most stably expressed (Saha and Vandemark 2012).

Recently, it has become feasible to produce transcriptomic resources for non-model species by next generation sequencing. Libraries comprising a total of 450 cDNA Mbp from flowers, leaves, cotyledons, epicotyl and hypocotyl, and light treated etiolated seedlings, were assembled into 324,428 unigenes and annotated in *A. thaliana*, *M. truncatula*, *G. max* and other databases (Franssen et al. 2011).

Comparative transcriptomics was used to investigate adaptation of *Rhizobium leguminosarum* to pea, sugar beet and alfalfa and has enabled differentiation between factors conserved across plants for rhizosphere colonization as well as identification of exquisite specific adaptation to host plants (Ramachandran et al. 2011). Transcriptome variations in reaction to abiotic and biotic stresses were also analyzed using several transcriptomic approaches. Currently, the microarray studies yielding more complex date are used to obtain a global view of gene expression and provide information about the possible mechanisms and pathways involved in the resistance (Fondevilla et al. 2011a). Similarly, chilling and acclimation mechanisms in freeze-tolerant pea line were compared with a sensitive line on transcriptome gene profiles and were associated with morphological measurements and histological observations (Lucau-Danilla et al. 2012). Further development in the microarray field led to other transcriptomic applications, such as detection of non-coding RNAs, single nucleotide polymorphisms (SNPs), and alternative splicing events. DNA sequencing approaches to transcriptome analysis have been an alternative to microarray-based methods.

Proteome analysis is becoming a powerful tool in the functional characterization of plants. With the availability of vast nucleotide sequence information and the progress achieved in sensitive and rapid protein identification by mass spectrometry, proteome approaches opened up new perspectives to analyze complex traits in plants at different levels. Subcellular compartments as mitochondria and chloroplast in pea were analyzed and proteins involve in specific function of these compartments were identified (Bardel et al. 2002; Taylor et al. 2005; Kanervo et al. 2008). Several proteome reference maps for leaves and stems were established. Studies of the developmental processes during germination were performed and the candidate proteins associated with the loss of desiccation tolerance were identified. Among the total of 139 protein spots showing a significant change during germination, number of new proteins were identified, such as sterol biosynthetic enzyme, ethylene biosynthetic enzyme, ACC oxidase, actin depolymerizing factor-like protein, the ROS detoxification enzyme, GPX, and actin reorganization factor. It was found, that seed germination involves not only in the activation of a series of metabolic processes, but also the reorganization of cellular structure and activation of protective systems (Wang et al. 2012). Finally, the pea mature seed proteome reference map containing high number of storage proteins was constructed, that give us new insights into the pea storage protein processing, especially in case of 7S globulins (Bourgeois et al. 2009). Combination of these results together with protein changes in their relative abundance during nitrogen mobilization from leaves to filling seeds, provide a complex view of cellular processes during germination (Schiltz et al. 2004). Several proteins involved in biotic and abiotic stress were also identified, including LEA, dehydrins and heat shock proteins (Grelet et al. 2005). Protein

differences in leaf proteome composition between pea genotypes displaying different phenotypes in response to pathogen infection were analyzed (Barilli et al. 2012; Castillejo et al. 2004). Relevant proteins were identified by mass spectrometry and their possible function was deduced. It was shown that most of the identified proteins corresponded to enzymes belonging to photosynthesis, metabolism, biosynthesis, binding and defense response pathways, with different behaviour patterns in relation to susceptibility/resistance of the studied genotypes. Similarly, root proteome variation associated with the root infection by a soil-borne pathogen was tested (Wen et al. 2007). In the pea–pathogen interaction, removal of border cells from roots prior to inoculation with spores resulted in a significant increase in frequency of root tip infection. These results support the hypothesis that proteins released along with the root cap mucilage during the process of border cell separation play a vital role in the plant's system of immunity, by protecting the root tip as it moves through the soil environment. Proteins are physically and chemically much more diverse than nucleic acids, which hinders the quantitative analysis of protein complexes. With the combination of proteomic and quantitative trait loci (QTL) mapping approaches, the genetic architecture of seed proteome variability was uncovered (Bourgeois et al. 2011a, b). Protein quantity loci (PQL) were searched for 525 spots detected on 2D-gels, most protein quantity loci was mapped in clusters. Interestingly, the analysis revealed that, the accumulation of the major storage protein families was under the control of a limited number of loci. To extend our knowledge of the pea genome structure, the current studies are focused on pea metabolome (Charlton et al. 2004, 2008). Recently, leaf metabolome has been profiled to monitor the changes induced by drought-stress and pea seed development (Charlton et al. 2008; Vigeolas et al. 2008).

Use of Genomic Resources in Molecular Breeding

Despite the effort and progress developing molecular resources, use of currently available in pea breeding has been limited. Several factors limit the direct application of QTLs and their associated markers including: (1) imprecise phenotypic description resulting in inaccurate marker-trait associations, (2) use of small mapping populations (50–200 individuals) resulting in limited genetic resolution, (3) lack of common markers reference markers across QTL studies, (4) limited range of variation in cultivated genepool, (5) trait and marker validation in different genetic background, (6) high genotype × environment interactions on expression, (7) necessity to test polymorphism of the molecular markers in different genetic backgrounds, (8) large (5–10 cM on average) genetic distances between markers and the QTLs, and finally small investments in pea and thus lag in molecular tool development for breeding (Smýkal et al. 2012a, b). However, efforts were made to use of genomic resources to molecular breeding of pea.

Mapping of Disease and Pest Resistance Genes in Pea

As in any other crop, there is long list of diseases and pests affecting pea. Among them fungal and viral pathogens is likely causing the most severe damage. Fusarium wilt is one of the most widespread diseases of pea, limiting pea production worldwide. Four races of *Fusarium oxysporum*, race 1, 2, 5 and 6, are recognized based on differential pathogenicity on pea cultivars (Kraft and Pfleger 2001). The infection occurs by wind, rain and environmental factors through movement of infested soil or seed. Field symptoms appear as patches of dead or dying plants, eventually becoming dry and brittle. Cultivation of resistant varieties offers the best solution to combat the disease (Kraft and Pfleger 2001). Available resistance to fusarium root rot caused by (*F. solani* f.sp. *pisi*) is quantitatively inherited, with three QTLs identified together with STMS markers for use in marker assisted breeding (Coyne et al. 2007a) (Table 15.2). Single genes are available for fusarium wilt resistance, such as *Fw* gene providing resistance to race 1, *Fwn* to race 2 or *Fwf* to race 5 placed on LGII (Kraft and Pfleger 2001; Grajal-Martin and Muehlbauer 2002; McClendon et al. 2002; Coyne et al. 2000; Okubara et al. 2005). Recently (Kwon et al. 2013) has developed user friendly markers linked to *Fusarium* wilt race 1, and provided its anchoring with mapped gene specific markers. AFLP and RAPD markers for *Fw* placed on LGIII were developed; however these have limited use for breeding owing to poor transferability (McClendon et al. 2002; Okubara et al. 2005). Putative single gene *Fnw* was recently placed on linkage group IV (LOD 40.0–65.6) and minor loci on LG III (LOD 3.97–4.60) (McPhee et al. 2012) (Table 15.2).

Ascochyta blight of pea is caused by a complex of three fungal pathogens: *Mycosphaerella pinodes*, *Phoma medicaginis* var. *pinodella* and *Ascochyta pisi*. Both single genes (*Rap2*) and QTLs have been reported conferring resistance to *A. pisi* (Darby et al. 1985). Similarly both single genes (*Rmp1*, *Rmp2*, *Rmp3* and *Rmp4*) and QTL have been described for *M. pinodes* resistance (Clulow et al. 1991; Timmerman-Vaughan et al. 2002, 2004; Tar'an et al. 2003; Prioul et al. 2004; Fondevilla et al. 2011b) (Table 15.2). Microarray technology was used to identify 346 differentially expressed genes in a resistant reaction to *Mycosphaerella pinodes* in pea among them genes involved in defense reactions such as cell wall reinforcement, phenylpropanoid and phytoalexins metabolism, pathogenesis-related proteins and detoxification processes (Fondevilla et al. 2011a). Three genes (*er1*, *er2* and *Er3*) have been postulated to confer resistance to powdery mildew (*Erysiphe pisi*) (Fondevilla et al. 2007; Heringa et al. 1969) (Table 15.2). The gene *er1* is in wide use in pea breeding programs (Heringa et al. 1969; Tiwari et al. 1997; Harland 1948). Resistance conferred by this gene has proven to be stable and is caused by a barrier to pathogen penetration (Fondevilla et al. 2006). Various markers have been linked to the *er1* locus mapped on linkage group VI (Ghafoor and McPhee 2012; Dirlewanger et al. 1994; Pereira et al. 2010; Srivastava et al. 2012; Ek et al. 2005). Recently, by study of novel *er1* allele, co-segregation with *PsMLO1* (Mildew Resistance Locus O) loss-of-function was reported (Pavan et al. 2011; Humphry et al. 2011). Analysis of the respective gene from several known powdery mildew

Table 15.2 Traits and identified QTL(s)

Trait	Mapping population	Environment	Marker types	R^2 or cM	Loci/QTL	Reference
Biotic stress resistance						
Ascochyta pisi resistance	Erygel × 661, 102 F2:3-4	Field, 1 environment	RFLP, RAPD	0.71	2 QTLs	Dirlewanger et al. (1994)
Erysiphe pisi resistance	955180 × Majoret, F2	Greenhouse	351 SSR	10.4–15 cM	*er1*	Ek et al. (2005)
Erysiphe pisi resistance	C2 × Messire, 100 F2, Pisum fulvum origin	Growth chamber	RADP/SCAR	2.8 cM	*Er3*	Fondevilla et al. (2008)
Erysiphe pisi resistance	Lincoln/JI2480,	Field, growth chamber	RADP/SCAR	2.6 cM	*er2*	Katoch et al. (2010)
Erysiphe pisi resistance	Fallon/11,760 3 F2, 117 genotypes	Field, 1 environment	RADP/SCAR	8.2 cM	*er1*	Nissar and Ghafoor (2011)
Erysiphe pisi resistance	2,200 M1	Growth chamber	CAPS	Gene	*MLO* locus, gene	Pavan et al. (2011)
Mycophaerella pinodes resistance	Erygel × 661, 102 F2:3-4	Field, 1 environment	RFLP, RAPD	0.71	3 QTLs	Dirlewanger et al. (1994)
Mycophaerella pinodes resistance	Carneval × MP1401, 88 RIL	Field, 11 environments	AFLP, SSR	0.08–0.14	3 QTLs	Tar'an et al. (2003)
Mycophaerella pinodes resistance	3148-A88 × Rovar, 133 F2:3-4	Field, 3 environments	AFLP, RAPD	0.09–0.35	13 QTLs	Timmerman-Vaughan et al. (2002)
Mycophaerella pinodes resistance	3147-A26 × Rovar, 148 F2:3-4	Field, 4 environments	AFLP, RAPD	0.07–0.21	11 QTLs	Timmerman-Vaughan et al. (2004)
Mycophaerella pinodes resistance	3148-A88 × Rovar, 133 F2:3-4	Field, 4 environments	AFLP, RAPD	0.07–0.36	14 QTLs	Timmerman-Vaughan et al. (2004)
Mycophaerella pinodes resistance	JI296 × DP, 135 RIL	Field and controlled, 1 environment	RAPD, SSR and STS	0.06–0.36	10 QTLs	Prioul et al. (2004)
Mycophaerella pinodes resistance	JI296 × DP, 135 RIL	Field and controlled, 1 environment	14 resistance gene analogs and defense genes	0.06–0.20	14 genes	Prioul-Gervais et al. (2007)
Mycophaerella pinodes resistance	JI296 × DP, 135 RIL	Growth chamber	RAPD, SSR and STS	0.05–0.18	6 QTLs	Prioul et al. (2004)
Mycophaerella pinodes resistance	Messire × P665 (P. syriacum), 111 RIL	Field and controlled, 1 environment	78 SSR	0.06–0.38	7 QTLs	Fondevilla et al. (2011b)
Aphanomyces euteiches resistance	Puget × 90-2079, 127 RIL	Field, 4 environments	AFLP, RAPDs, SSRs, ISSRs, STSs, isozymes	0.06–0.47	7 QTLs	Pilet-Nayel et al. (2002)
Aphanomyces euteiches resistance	Puget × 90-2079, 127 RIL	Greenhouse	AFLP, RAPDs, SSRs, ISSRs, STSs, isozymes	0.08–0.16	6 QTLs	Pilet-Nayel et al. (2005)
Aphanomyces euteiches resistance	Baccara × 552, Baccara × PI 180693 178, 178 RIL	Multiple	SSR, RAPD, RGA	0.06–0.27	135 QTLs	Hamon et al. (2010)

Fusarium solani resistance	CMG × PI220174, 275 RIL	Greenhouse	SSR, RAPD, RGA	–	3 QTLs	Weden and Porter (2007)
Fusarium solani resistance	Carman × Reward, 71 RIL	Field, greenhouse	SSR	0.39	1 QTL	Feng et al. (2011)
Fusarium oxysporum race 1 resistance	Green Arrow × PI 179449, 80 RIL	Field, greenhouse	AFLP, RAPD into SCAR	4 and 4.6 cM	*Fw*	Okubara et al. (2005)
Fusarium oxysporum race 2 resistance	Shawnee × Bohaty, 187 RIL	Field, 3 environments	176 SSR, 88 RAPD, five isozymes	0.028–0.80	3 QTLs	McPhee et al. (2012)
Orobanche crenata resistance	Messire × P665 (P. syriacum), 115 F3	Field, 1 environment	RAPD, SSR	0.10–0.11	2 QTLs	Valderrama et al. (2004)
Orobanche crenata resistance	Messire × P665 (P. syriacum), 111 RIL	Field, 2 environment	RAPD, SSR	0.10–0.17	4 QTLs	Fondevilla et al. (2009)
Orobanche crenata resistance	Messire × P665 (P. syriacum), 111 RIL	Growth chamber	RAPD, SSR	0.13–0.37	7 QTLs	Fondevilla et al. (2009)
O. crenata field resistance	Messire × P665 (P. syriacum), 111 RIL	Field	78 SSR	0.08–0.39	5 QTLs	Fondevilla et al. (2011b)
O. crenata seed germination	Messire × P665 (P. syriacum), 111 RIL	Field, growth chamber	78 SSR	0.08–0.39	2 QTLs	Fondevilla et al. (2011b)
O. crenata tubercle development	Messire × P665 (P. syriacum), 111 RIL	Field, growth chamber	78 SSR	0.08–0.39	3 QTLs	Fondevilla et al. (2011b)
Uromyces pisi resistance	94 RIL		RAPD, SSR	0.60	1 QTL	Barilli et al. (2010)
Uromyces fabae resistance	HUPV × (HUVP 1 × FC 1) 31 BC1F1	Field, greenhouse	614 RAPD	10.8 and 24.5 cM	1 QTL	Vijayalakshmi et al. (2005)
Uromyces fabae resistance	HUVP 1 × FC 1, 136 RIL	Field, 4 environment, greenhouse	148 SSR, RAPD, STS	0.22 and 0.58	2 QTLs	Rai et al. (2011)
Pseudomonas syringae resistance	Vinco × Hurst's Greenshaft, Partridge × Early Onward F2, JI281 × JI399 RILs	Greenhouse	11 isozymes	–	4 QTLs	Hunter et al. (2001)
Pseudomonas syringae resistance	Messire × P665 (P. syriacum), 111 RIL	Greenhouse	RAPD, SSR	0.08–0.22	2 QTLs	Fondevilla et al. (2012)
Pea Seedborne Mosaic Virus (PSbMV) resistance	88 V1.11 × 425, 79 F2, Primo × OSU442-15, 47 F2	Greenhouse	RAPD, AFLP	4 cM	*sbm-1*	Frew et al. (2002)
Pea Seedborne Mosaic Virus (PSbMV) resistance	JI1405 × Scout, 500 F2	Greenhouse	cDNA-AFLP	0.7 cM	*eIF4E*	Gao et al. (2004)
Bruchus resistance	Alaska 81 × Pisum fulvum (PI 595946, PI 343955), F2:3	Greenhouse	phenotype assessment	–	–	Clement et al. (2009)
Abiotic stress resistance						
Winter Frost tolerance	Champagne × Terese, 164 RIL	Field, 11 environments	213 SSR and genes	0.03–0.49	6 QTLs	Lejeune-Hénaut et al. (2008)
Winter Frost tolerance	Champagne × Terese, 78 RIL	Field, 2 environments	SSR and biochemical	0.06–0.46	4 QTLs	Dumont et al. (2009)

(continued)

Table 15.2 (continued)

Trait	Mapping population	Environment	Marker types	R^2 or cM	Loci/QTL	Reference
Frost Damage	Champagne × Terese, 78 RIL	Growth chamber	SSR and biochemical	0.11–0.30	4 QTLs	Dumont et al. (2009)
Raffinose content	Champagne × Terese, 78 RIL	Field and growth chamber (3 environs)	SSR and biochemical	0.09–0.25	2 QTLs	Dumont et al. (2009)
Glucose content	Champagne × Terese, 78 RIL	Field and growth chamber (3 environs)	SSR and biochemical	0.12	1 QTL	Dumont et al. (2009)
Electrolyte leakage	Champagne × Terese, 78 RIL	Field and growth chamber (3 environs)	SSR and biochemical	0.16	1 QTL	Dumont et al. (2009)
Rubisco content	Champagne × Terese, 78 RIL	Field and growth chamber (3 environs)	SSR and biochemical	0.03–0.12	1 QTL	Dumont et al. (2009)
Seed traits						
Seed weight	Primo × OSU 442-15, 105 F2	Field, 1 environment	RAPD, AFLP	0.13–0.26	3 QTLs	Timmerman-Vaughan et al. (1996)
Seed weight	JI1794 × Slow, 51 RIL	Greenhouse	RAPD, AFLP	–	7 QTLs	Timmerman-Vaughan et al. (1996)
Seed weight	WT10245 × Wt11238	Field, 1 environment	AFLP, ISSR, STS, CAPS and RAPD	0.10–0.24	5 QTLs	Irzykowska and Wolko (2004)
Seed weight	Primo × OSU 442-15, 227 F2:3-4	Field, 1 environments	SSR, RAPD	0.03–0.19	10 QTLs	Timmerman-Vaughan et al. (2005)
Seed weight	Terese × K586, 139 RIL	Field, 5 environments	SSR, gene-based	0.09–0.32	9 QTLs	Burstin et al. (2007)
Seed yield	Terese × K586, 139 RIL	Field, 5 environments	SSR, gene-based	0.09–0.53	6 QTLs	Burstin et al. (2007)
Seed yield	WT10245 × Wt11238	Field, 1 environment	AFLP, ISSR, STS, CAPS and RAPD	0.09–0.47	6 QTLs	Irzykowska and Wolko (2004)
Seed yield	Carneval × MP1401, 88 RIL	Field, 13 environment	AFLP, STS, RAPD	0.39	4 QTLs	Tar'an et al. (2005)
Seed yield	Primo × OSU 442-15, 227 F2:3-4	Field, 1 environments	RAPD, SSR	0.05–0.15	5 QTLs	Timmerman-Vaughan et al. (2005)
Seed yield	Cameor × Ballet, 207 RIL plus 7 genotypes	Field, 2 environments	SSR, gene-based		4–2 QTLs	Bourion et al. (2010)
Yield components	Wt11238 9 Wt3557, Wt10245 9 Wt11238, 110, 101 RIL	Multiple	AFLP, RAPD, STS, CAPS, ISSR	–	77 QTLs	Krajewski et al. (2012)
Seed number	Primo × OSU 442-15, 227 F2:3-4	Field, 1 environments	RAPD, SSR	0.03–0.27	9 QTLs	Timmerman-Vaughan et al. (2005)

Seed number	WT10245 × Wt11238	Field, 1 environment	AFLP, ISSR, STS, CAPS and RAPD	0.10–0.55	5 QTLs	Irzykowska and Wolko (2004)
Seed number	Terese × K586, 139 RIL	Field, 5 environments	SSR, gene-based	0.10–0.44	5 QTLs	Burstin et al. (2007)
Seed number	Cameor × Ballet, 207 RIL plus 7 genotypes	Field, 2 environments	SSR, gene-based	0.20–0.55	2–3 QTLs	Bourion et al. (2010)
Protein content	WT10245 × Wt11238	Field, 1 environment	AFLP, ISSR, STS, CAPS and RAPD	0.13–0.25	5 QTLs	Irzykowska and Wolko (2004)
Protein content	Carneval × MP1401, 88 RIL	Field, 13 environments	AFLP, STS, RAPD	0.45	3 QTLs	Tar'an et al. (2005)
Protein content	Terese × K586, 139 RIL	Field, 5 environments	SSR, gene-based	0.10–0.46	8 QTLs	Burstin et al. (2007)
Protein content	Cameor × Ballet, 207 RIL plus 7 genotypes	Field, 2 environments	SSR, gene-based	0.09–0.21	2–4 QTLs	Bourion et al. (2010)
Protein content	Cameor × VavD265, Cameor × Ballet, 157, 180RIL	Field, 1 environment	525 PQL (protein quantity loci)	–	2–6 QTLs	Bourgeois et al. (2011a, b)
Protein content	Cameor × VavD265, Cameor × Ballet, 157, 180RIL	Field, 1 environment	525 PQL (protein quantity loci)	–	2–4 QTLs	Bourgeois et al. (2011a, b)
Protein content	Cameor × VavD265, Cameor × Ballet, 157, 180RIL	Field, 1 environment	525 PQL (protein quantity loci)	–	5 QTLs	Bourgeois et al. (2011a, b)
Protein digestibility	Cameor × VavD265, Cameor × Ballet, 157, 180RIL	Field, 1 environment	525 PQL (protein quantity loci)	–	1 QTLs	Bourgeois et al. (2011a, b)
245 polypeptides' quantity (2D-PAGE)	Cameor × VavD265, Cameor × Ballet, 157, 180RIL	Field, 3 environments	525 PQL (protein quantity loci)	–	1–6 QTLs	Bourgeois et al. (2011a, b)
Number of cotyledon cells	Terese × K586, 139 RIL	Field, 5 environments	SSR, gene-based	0.19–0.27	2 QTLs	Burstin et al. (2007)
Volume of cotyledon cells	Terese × K586, 139 RIL	Field, 5 environments	SSR, gene-based	0.20–0.32	4 QTLs	Burstin et al. (2007)
Yellow seed lightness	Alfetta × CDC Bronco, 92 RIL, Orb × CDC Striker 120 RIL	Field, 2 environments	AFLP, SSR	0.08–0.19	9 QTLs	Ubayasena et al. (2011)
Yellow seed greeness	Alfetta × CDC Bronco, 92 RIL, Orb × CDC Striker 120 RIL	Field, 2 environments	AFLP, SSR	0.11–0.26	3 QTLs	Ubayasena et al. (2011)
Seed shape	Alfetta × CDC Bronco, 92 RIL, Orb × CDC Striker 120 RIL	Field, 2 environments	AFLP, SSR	0.25–0.28	15 QTLs	Ubayasena et al. (2011)
Seed dimpling	Alfetta × CDC Bronco, 92 RIL, Orb × CDC Striker 120 RIL	Field, 2 environments	AFLP, SSR	0.13–0.23	9 QTLs	Ubayasena et al. (2011)

(continued)

Table 15.2 (continued)

Trait	Mapping population	Environment	Marker types	R^2 or cM	Loci/QTL	Reference
Loss of hillum, Def locus	"DGV" (def wild-type) and "PF" (def mutant) 472 F2	F2-3 assessment	AFLP	–	1–5 cM from *def*	von Stackelberg et al. (2003)
Plant phenology						
Days to maturity	Carneval × MP1401, 88 RIL	Field, 13 environments	AFLP, SSR	0.35	4 QTLs	Tar'an et al. (2005)
First flowering node	Primo × OSU 442-15, 227 F2:3-4	Field, 3 environments	AFLP, RAPD, SSR	0.06–0.20	4 QTLs	Timmerman-Vaughan et al. (2005)
Beginning of flowering	Terese × K586, 139 RIL	Field, 5 environments	SSR	0.12–0.67	8 QTLs	Burstin et al. (2007)
End of flowering	Terese × K586, 139 RIL	Field, 5 environments	SSR	0.11–0.47	8 QTLs	Burstin et al. (2007)
50 % flowering	Messire × P665 (P. syriacum), 111 RIL	Field, 1 environment	SSR		4 QTLs	Fondevilla et al. (2011b)
Flowering time locus Hr	Champagne × Terese, 164 RIL	Field, 11 environments	SSR	0.03–0.49	6 QTLs	Lejeune-Hénaut et al. (2008)
Photoperiod response	Multiple RIL	Multiple	Multiple, some identified as gene	–	8 genes	Weller et al. (2009)
Circadian clock	Multiple RIL	Multiple	Multiple, some identified as gene	–	4 genes	Weller et al. (2009)
Inflorescence identity loci	Multiple RIL	Multiple	Multiple, some identified as gene	–	8 genes	Weller et al. (2009)
Stem length	L2 × L77,205 isoline F1:3 families	–	–	–	*Le* gene	Lester et al. (1997)
Plant morphology						
Lodging resistance	Carneval × MP1401, 88 RIL	Field, 11 environments	AFLP, SSR	0.18–0.51	2 QTLs	Tar'an et al. (2003)
Plant height	Carneval × MP1401, 88 RIL	Field, 11 environments	AFLP, SSR	0.06–0.47	3 QTLs	Tar'an et al. (2003)
Plant height	Terese × K586, 139 RIL	Field, 5 environments	SSR, gene-based	0.09–0.63	4 QTLs	Burstin et al. (2007)
Root length	Messire × P665 (P. syriacum), 111 RIL	Field and controlled, 1 environment	78 SSR	0.06–0.38	3 QTLs	Fondevilla et al. (2011b)
Total root length	Cameor × Ballet, 207 RIL plus 7 genotypes	Field, 2 environments	SSR, gene-based	0.1–0.45	1–2 QTLs	Bourion et al. (2010)
Total root elongation rate	Cameor × Ballet, 207 RIL plus 7 genotypes	Field, 2 environments	SSR, gene-based	0.1–0.33	4 QTLs	Bourion et al. (2010)

Root biomass	Cameor × Ballet, 207 RIL plus 7 genotypes	Field, 2 environments	SSR, gene-based	0.09–0.49	32 QTLs	Bourion et al. (2010)
Root biomass	JI1794 × Slow, 51 RIL	Greenhouse	RAPD, STS	–	associated to *Le*	Weeden and Moffet (2002)
Number of lateral root	Cameor × Ballet, 207 RIL plus 7 genotypes	Field, 2 environments	SSR, gene-based	0.1–0.2	2–4 QTLs	Bourion et al. (2010)
Nodule biomass	Cameor × Ballet, 207 RIL plus 7 genotypes	Field, 2 environments	SSR, gene-based	0.09–0.35	8 QTLs	Bourion et al. (2010)
Number of nodules	Cameor × Ballet, 207 RIL plus 7 genotypes	Field, 2 environments	SSR, gene-based	not given	9 QTLs	Bourion et al. (2010)
Plant biomass	Cameor × Ballet, 207 RIL plus 7 genotypes	Field, 2 environments	SSR, gene-based	not given	2–4 QTLs	Bourion et al. (2010)
Plant biomass	Terese × K586, 139 RIL	Field, 5 environments	SSR, gene-based	0.09–0.64	6 QTLs	Burstin et al. (2007)
Vegetative biomass at harvest	Terese × K586, 139 RIL	Field, 5 environments	SSR, gene-based	0.09–0.64	1–5 QTLs	Burstin et al. (2007)
Vegetative biomass at harvest	Cameor × Ballet, 207 RIL plus 7 genotypes	Field, 2 environments	SSR, gene-based	not given	7 QTLs	Bourion et al. (2010)
Number of basal branches	Terese × K586, 139 RIL	Field, 5 environments	SSR, gene-based	0.09–0.64	3–4 QTLs	Burstin et al. (2007)
Number of basal branches	Terese × K586, 139 RIL	Field, 5 environments	SSR, gene-based	0.09–0.64	7 QTLs	Burstin et al. (2007)
Number of nodes	Primo × OSU 442-15, 227 F2:3-4	Field, 3 environments	RAPD, AFLP	0.06–0.33	6 QTLs	Timmerman-Vaughan et al. (2005)
Number of flowering nodes	Primo × OSU 442-15, 227 F2:3-4	Field, 3 environments	RAPD, AFLP	0.07–0.12	5 QTLs	Timmerman-Vaughan et al. (2005)
Harvest index	Primo × OSU 442-15, 227 F2:3-4	Field, 3 environments	RAPD, AFLP	0.06–0.20	4 QTLs	Timmerman-Vaughan et al. (2005)
Harvest index	Terese × K586, 139 RIL	Field, 5 environments	SSR, gene-based	0.11–0.66	7 QTLs	Burstin et al. (2007)
Plant physiology						
Nitrogen harvest index	Terese × K586, 139 RIL	Field, 5 environments	SSR, gene-based	0.09–0.18	8 QTLs	Burstin et al. (2007)
Nitrogen nutrition index	Terese × K586, 139 RIL	Field, 5 environments	SSR, gene-based	0.1–0.25	3 QTLs	Burstin et al. (2007)
Plant vegetative biomass at harvest	Terese × K586, 139 RIL	Field, 5 environments	SSR, gene-based	0.15–0.35	7 QTLs	Burstin et al. (2007)
Plant %N at BSF	Terese × K586, 139 RIL	Field, 5 environments	SSR, gene-based	0.1–0.31	6 QTLs	Burstin et al. (2007)
Plant %N at harvest	Terese × K586, 139 RIL	Field, 5 environments	SSR, gene-based	0.11–0.17	5 QTLs	Burstin et al. (2007)
Seed nitrogen content, plant productivity	Terese × K586, 139 RIL	Field, 5 environments	SSR, gene-based	0.08–0.31	261 QTLs	Burstin et al. (2007)

(continued)

Table 15.2 (continued)

Trait	Mapping population	Environment	Marker types	R^2 or cM	Loci/QTL	Reference
N acquisition eficiency	Cameor × Ballet, 207 RIL plus 7 genotypes	Field, 2 environments	SSR, gene-based	0.09–0.21	10	Bourion et al. (2010)
N accumulation	Cameor × Ballet, 207 RIL plus 7 genotypes	Field, 2 environments	SSR, gene-based	0.09–0.21	7	Bourion et al. (2010)
C accumulation	Cameor × Ballet, 207 RIL plus 7 genotypes	Field, 2 environments	SSR, gene-based	0.09–0.21	26	Bourion et al. (2010)
Association mapping						
Seed micronutrients	285 accessions, USDA core	49 phenotypic traits	137 DNA markers, 15 SSR, 36 RAPD, 1 SCAR	0.02–0.061	10 markers	Kwon et al. (2012)
Disease/pest resistances	285 accessions, USDA core	49 phenotypic traits	137 DNA markers, 15 SSR, 36 RAPD, 1 SCAR	0.012–0.245	11 markers	Kwon et al. (2012)
Morphology/agronomy	285 accessions, USDA core	49 phenotypic traits	137 DNA markers, 15 SSR, 36 RAPD, 1 SCAR	0.019–0.058	42 markers	Kwon et al. (2012)
Phenology	285 accessions, USDA core	49 phenotypic traits	137 DNA markers, 15 SSR, 36 RAPD, 1 SCAR	0.024–0.225	4 markers	Kwon et al. (2012)
Production	285 accessions, USDA core	49 phenotypic traits	137 DNA markers, 15 SSR, 36 RAPD, 1 SCAR	0.021–0.037	10 markers	Kwon et al. (2012)

resistant cultivars has further supported that indeed *PsMLO1* loss-of-function is responsible for the trait and indicated the same molecular basis are shared among well studied barley *mlo*, tomato *ol-2* and pea *er1*genes (Pavan et al. 2011; Humphry et al. 2011). Gene *er2* mapped on linkage group III confers a high level of resistance only in some locations as its expression is strongly influenced by temperature and leaf age (Fondevilla et al. 2006; Katoch et al. 2010; Tiwari et al. 1993). Novel, dominant resistance gene *Er3* was identified recently in *P. fulvum* and has been introduced successfully into adapted *P. sativum* material by sexual crossing (Fondevilla et al. 2007, 2008). Resistance conferred by *Er3* is due to a high frequency of cell death that occurs both as a rapid response to attempted infection and a delayed response that follows colony establishment (Fondevilla et al. 2007). *Erysiphe pisi* is distributed worldwide and there is someevidence of physiological specialization and resistance to disease conferred by er1 was overcome by anisolate of the fungus obtained from naturally infected plants. Importantly, *Erysiphe baumleri* and *Erysiphe trifolii* have been recently reported that was not previously known as a pathogen of pea powdery mildew (Ondřej et al. 2005; Attanayake et al. 2010).

Pea rust has been reported to be caused by *Uromyces viciae-fabae* in tropical and subtropical regions and or by *U. pisi* in temperate regions (Kushwaha et al. 2006; Barilli et al. 2010). Both major genes (*Ruf*) and QTLs have been reported conferring resistance to *U. viciae-fabae* in pea (Vijayalakshmi et al. 2005; Rai et al. 2011) (Table 15.2). Mapping revealed one major and one minor (environment specific) QTL for rust resistance on LGVII, flanked by SSR markers at 10.8 cM, explaining 42.4 and 58.8 % of the total phenotypic variation (Barilli et al. 2010). Genetic resistance to *Aphanomyces* root rot (*Aphanomyces euteiches*) in pea is reported to be governed either by a single gene or by QTLs (Weeden et al. 2000; Pilet-Nayel et al. 2002, 2005). Two mapping populations of 178 recombinant inbred lines each were used to identify QTL for *Aphanomyces* root rot resistance in controlled and in multiple French and USA field conditions. This led to identification of total of 135 additive-effect QTL corresponding to 23 genomic regions and 13 significant epistatic interactions associated with partial resistance to *A. euteiches* in pea (Hamon et al. 2010) (Table 15.2). Using a linkage map based on RAPD, SSR and STS marker polymorphism of a RIL population, identified 10 and 6 QTLs associated with Ascochyta blight resistance at the seedling stage and adult plant stages, respectively, and four more developmental stage independent QTLs (Pilet-Nayel et al. 2005). In addition to Ascochyta blight resistance, three QTLs for flowering date and plant height were reported. *Pseudomonas syringae* is the main pathogen responsible for bacterial blight in pea and can cause yield losses of 70 % in cool and wet conditions. Resistance to pea bacterial blight (*Pseudomonas syringae pv. pisi*) is controlled by single dominant genes *Ppi1* (for race R2), *Ppi3* (R3) and *Ppi4* (R4) (Hunter et al. 2001). Recently, two QTLs associated with resistance, *Psy1* and *Psy2*, were identified, explaining 22.2 and 8.6 % of the phenotypic variation, respectively (Fondevilla et al. 2012) (Table 15.2). *P. abyssinicum* accessions (16 originated from Ethiopia and 1 from Yemen) were identified to be resistant or partially resistant to all races including race 6, for which there are no known commercial resistant cultivars. This resistance is controlled by a major recessive gene together with a number of modifiers (Elvira-Recuenco and Taylor 2001).

Incomplete resistance to parasitic plant, crenate broomrape (*Orobanche crenata*) is available in accessions of *P. sativum* ssp. *sativum*, *abyssinicum*, *arvense* and *elatius* and in *P. fulvum* and these are being introgressed into cultivated pea (Rubiales et al. 2005, 2009). Resistance to pea bruchid (*Bruchus pisorum*) has been described in *P. fulvum* being conferred by three genes (Byrne et al. 2008; Clement et al. 2009). Pea lines with the *Np* gene respond to the presence of pea weevil eggs on pods by forming callus (neoplastic pod trait) that reduces larval entry into the pod (Hardie et al. 1995; Doss et al. 2000). In a field trial, this pod-based resistance was responsible for a lower rate of weevil infested seed (62.2 %) in *Np* plants compared to that in a susceptible line (85.4 %) (Doss et al. 1995). Recently using model species of *Medicago truncatula* and *Acyrthosiphon pisum* aphid has led to identification of major QTLs associated with defense which might be expected to be transferable to pea (Guo et al. 2012). In addition to fungal, virus diseases are among the most widespread and destructive pathogens of crop plants causing serious economic losses by yield and quality reduction. Genetics of potyviruses resistance has been thoroughly studied in pea, defining one cluster (on linkage group II) of *bcm*, *cyv-1*, *mo*, *pmv* and *sbm-2loci* conferring resistance to Bean common mosaic virus (BCMV), Clover yellow vein virus (ClYVV), Pea mosaic virus (PMV) and L1 (P2) pathotype of Pea seed-borne mosaic virus (PSbMV) (Provvidenti and Hampton 1991). The second cluster (on linkage group VI) includes *cyv-2*, *wlv* and *sbm-1*, conferring resistance to Clover yellow vein virus (ClYVV), white lupin strain of Bean yellow mosaic virus (BYMV-W) and P1 pathotype of PSbMV (Provvidenti and Hampton 1991). The *eIF4E* gene was identified by candidate gene approach to be responsible for PSbMV, BYMV-W, ClYVV resistances at *sbm-1*/*cyv-2*locus (Gao et al. 2004; Bruun-Rasmussen et al. 2007; Johansen et al. 2001; Andrade et al. 2009). Based on these studies, reliable and allele specific testing was developed (Smýkal et al. 2010). Another destructive and widely spread Pea enation mosaic virus, from the family Luteoviridae is transmitted by aphids in a non-propagative, circulative and persistent mode. Resistance to PEMV was found to be quantitative although recently the gene for resistance to *En*, is closely linked to *Prx1* with an estimated recombination frequency of 0.02 was reported (Timmerman-Vaughan et al. 2009) (Table 15.2).

Mapping of Agronomic Traits

Compared to other economically important crops, less QTL mapping studies for agronomical traits have been reported in pea (McPhee 2007; Swiecicki and Timmerman-Vaughan 2005; Burstin et al. 2007). First QTL analysis in pea generated a genetic linkage map of two populations segregating for seed weight using 101 RFLP, 58 RAPD and 40 AFLP markers (Timmerman-Vaughan et al. 1996). One of the major QTLs identified on LG III was mapped to orthologous regions responsible for control of seed weight in *Vigna* and soybean (Timmerman-Vaughan et al. 1996) (Table 15.2). Another genetic linkage maps based on AFLP, RAPD and STS markers and reported QTLs associated with lodging resistance (QTLs on LG III and VI), plant height (QTLs on LG III and two on unassigned LGs C and D), mycosphaerella blight (QTLs on LG II, IV and VI), grain yield (QTLs on LG II, VI and VII), seed

protein concentration (QTLs on LG III, VI and unassigned LG A), and maturity (QTLs on LG II, III and VI) (Tar'an et al. 2003, 2005) (Table 15.2). Lodging resistance markers developed in these studies were verified and applied in breeding program (Zhang et al. 2006) (Table 15.2). Genetic control of green cotyledon color in field pea and associated QTLs was studied and several QTLs associated with cotyledon color on LG III, IV, V and VII were reported (McCallum et al. 1997) (Table 15.2). QTLs associated with node number (3), earliness (2), plant height (1) and resistance to Ascochyta blight (1) also detected (Dirlewanger et al. 1994) (Table 15.2). Genetic control of cotyledon bleaching resistance was studied, using RILs which were phenotyped in field trials over four station-years and genotyped (Ubayasena et al. 2010) (Table 15.2). Heritability estimates for whole seed and cotyledon greenness were moderate (0.72 and 0.69, respectively), and increased when assessed after exposing whole seeds to accelerated bleaching conditions. Multiple QTL mapping (MQM) detected major QTLs on LGIV and LGV, as well as location- and year-specific QTLs on LGII and LGIII associated with green cotyledon bleaching resistance (Ubayasena et al. 2010) (Table 15.2). Same authors identified QTLs associated with visual quality of field pea including seed coat color, seed shape and seed dimpling (Ubayasena et al. 2011) (Table 15.2). Nine QTLs controlling yellow seed lightness, 3 for yellow seed greenness, 15 for seed shape and 9 for seed dimpling were detected. Among them, five QTLs located on LG II, LG IV and LG VII were consistent in at least 2 years (Ubayasena et al. 2011). A quantitative trait loci (QTL) approach was used to identify chromosomal regions linked to frost tolerance, explaining from 6.5 to 46.5 % of the phenotypic variance (Dumont et al. 2009) (Table 15.2). Flowering time is being studied by comparative candidate approach. Wild *P. sativum* ssp. *elatius* and a subset of pea landraces and winter cultivars do not flower under short photoperiods, but this long-day requirement has been genetically relaxed in a majority of cultivated lines. Up to six loci contribute to "natural" variation related to flowering in pea, with derived or cultivated alleles generally conferring early flowering and a reduction in photoperiod response. In addition, numerous other loci have been identified through mutational studies (Weller et al. 2009). Although there are difficulties to compare among QTL studies, general conclusions are consistent (e.g. Prioul et al. 2004; Weeden 2007) identifying QTL on LGII and LGIII in positions roughly corresponding to known positions of the *Lf* and *Hr* loci (Murfet 1973, 1975) (Prioul et al. 2004; Murfet 1973, 1975; Weeden 2007). *Lf* was the first pea flowering locus to be cloned, and was identified as a homolog of the *Arabidopsis* inflorescence identity gene *TFL1* (Foucher et al. 2003). Another *TFL1* homolog, *Det*, controls determinacy of the primary inflorescence in several legumes including pea, soybean and bean (Foucher et al. 2003). A "functional candidate" approach has also been used to clone the photoperiod response locus *Hr*, a major locus controlling flowering time, with recessive hr alleles causing reduction but not complete loss of the response to photoperiod (Murfet 1973). Comparative mapping has identified *Hr* as the pea ortholog of *Arabidopsis ELF3* (Weller et al. 2012). A single functional variant is widespread in pea germplasm and likely to underlie many of the flowering time QTL identified in this region of LGIII. Naturally-occurring recessive alleles at the *Sn* locus confer early flowering and completely eliminate the photoperiod response, but have a

restricted distribution within cultivated pea germplasm and may have arisen within a spring (*hr*) background. Like *Hr*, the *Sn* locus also appears to control circadian rhythms, and has also recently been identified as the pea ortholog of an *Arabidopsis* circadian clock gene (Hecht et al. 2007; Liew 2011). The flowering locus *Hr* was implicated to influence winter frost tolerance by delaying floral initiation until after the main winter freezing periods have passed (Lejeune-Hénaut et al. 2008). The dominant allele of *Hr* locus was found in a set of forage cultivars, which remain vegetative until a threshold day length of 13 h30 is reached. Although the underlaying gene was not yet cloned, identified three consistent QTLs: WFD 3.1/*Hr*, WFD 5.1/*Tri* and WFD 6.1/*Le* makes these loci interesting targets for marker assisted selection (Lejeune-Hénaut et al. 2008). Moreover, the flowering allele *Hr* enhances the capacity of pea photoperiodic lines to produce basal laterals, which is often found in primitive accessions. Agronomically, the recessive *le* allele is required in the dry pea cultivars for northern Europe, where all the dry peas are dwarf to minimize crop lodging. The *Tri* locus/WFD 5.1 QTL needs consideration in breeding programs, as the favorable allele for WFD 5.1 could bring together the dominant *Tri* allele which is unfavorable for a seed use for animal feeding, because it contains two structural genes encoding the major pea seed trypsin inhibitors. Set of PCR primers suitable for breeding for low trypsin inhibitor activity was developed (Page et al. 2002).

Genetic studies have identified a minimum of three QTL associated with total seed nitrogen and protein content (Bastin et al. 2007; Tar'an et al. 2005; Irzykowska and Wolko (2004)) (Table 15.2). QTL for seed yield and seed yield components in five different environments were mapped: 261 QTL were detected across the five environments for all traits measured (Bastin et al. 2007) (Table 15.2). Most QTL for seed traits mapped in clusters with plant traits, suggesting the significant role of source-sink interactions in the control of seed traits. Developmental genes *Le* and *Afila*, which control internode length and the switch between leaflets and tendrils, respectively, determined seed protein content and/or yield depending on the environment (Mikić et al. 2011) (Table 15.2). However not only quantity but also quality of seed proteins is important, as shown by pea albumin 2 (PA2) and lectin, presence of both negatively correlating with digestability (Chinoy et al. 2011). The pea seed protein composition was deciphered through a PQL approach (Bourgeois et al. 2011a, b). These authors mapped the loci controlling the quantity of 525 protein spots revealed by 2D-PAGE and found that the accumulation of the major storage protein families was under the control of a limited number of loci. Storage protein accumulation was under the control of both cis- and trans-regulatory regions. A locus on LGII appeared a major regulator of protein composition and of protein in vitro digestibility.

Pea Transgenesis and Mutagenesis

Although pea is accessible to genetic transformation, this remains a challenge and precludes systematic characterization of gene functions (Somers et al. 2003; Svabova et al. 2005). This is both because of recalcitrance of pea as most of legume species to

in vitro regeneration as well as *Agrobacterium*-mediated transformation (Atif et al. 2013). In spite of this, co-cultivation process was elaborated and several successful pea transformations were published (Atif et al. 2013; Svabova and Griga 2008). In addition to Agrobacterium-mediated, direct gene transfer methods such as electroporation of isolated pea protoplasts (Puonti-Kaerlas et al. 1999) and biolistic (Molnár et al. 1999; Warkentin et al. 1992). Recent review of legumes transformation has summarized successful and published pea transgenosis (Atif et al. 2013). Despite the fact that pea transformation was reported over 20 years ago its efficiency remains low (in range of 0.1–6.5 %) (Atif et al. 2013). Majority of these studies used only selection and reporter marker genes, but some used agronomically useful genes such as bean alpha-amylase inhibitor, tested even in field conditions and found effective against pea weevil (Schroeder et al. 1993, 1994; Morton et al. 2000). Unfortunately, later bean alpha-amylase inhibitor transgenic peas were found to have altered structure and immunogenicity (Prescott et al. 2005). Transgenic approach against PSbMV and PEMV viruses were successfully tested (Chowrira et al. 1998). Recently, the issue of transgenic pea came into focus in relation to plant-made vaccines, to which legumes; protein rich seeds are very suitable. Thus pea seeds expressing vaccines against rabbit haemorrhagic disease virus, intestinal infections in pigs, coccidiosis in chickens or tumor-associated carcinoembryonic antigen (Mikschofsky et al. 2009; Zimmerman et al. 2009; Perrin et al. 2000; Saalbach et al. 2001). Recently, virus-induced gene silencing (VIGS) has become an important reverse genetics tool for functional genomics and VIGS vectors based on Pea early browning virus (PEBV, genus Tobravirus) are available and were successfully used to silence pea genes involved in the symbiosis with nitrogen-fixing *Rhizobium* as well as development (Gronlund et al. 2010; Constantin et al. 2004). Virus mediated transgenosis was used to produce human acidic fibroblast growth factor in pea (Fann et al. 2011).

RNA interference (RNAi) mechanism, post-transcriptional process triggered by the introduction of double-stranded RNA (dsRNA) which leads to gene silencing in a sequence-specific manner, was used in the function study of biosynthetic pathways genes and the transcript level of the lotus Clv2 gene in pea (Kaimoyo and Van Etten 2008; Krusell et al. 2008). The genomics tools such as fast neutron and TILLING mutant populations were developed for reverse genetics approaches (Dalmais et al. 2008; Wang et al. 2008). The TILLING (targeting-induced local lesions in genomes) method combines the induction of a high number of random point mutations with mutagens like ethyl methane sulfonate (EMS) and mutational screening systems to discover induced mutations in sequence DNA targets. Sufficiently large TILLING population made in variety Cameor is available for pea and data were developed into on-line database, UTILLdb, that contains phenotypic as well as sequence information on mutant genes (Dalmais et al. 2008). Currently it has 4,817 lines, of them 1,840 with phenotype and 464 identified mutations by sequencing. Moreover, the commercial pea variety Cameor used for TILLING population has also been used for BAC library development, essential tool for positional cloning and also for pea genome sequencing (Hellens et al. 2010). Another BAC library was developed from PI269818 accession, used to introgress genetic diversity into the cultivated germplasm pool, which could be useful for the isolation of genes underlying disease resistance (such as *Fw*, Fusarium resistance loci) and other

economically important traits (Coyne et al. 2007b). Both BAC libraries would be essential for good quality pea genome sequencing.

Progress Towards Pea Genome Sequencing

Pea has been important cool season food legume, it is losing competitiveness because it does not have sufficient genomic resources as currently available to other crops. As a close relative of soybean, chickpea, cowpea, common bean, peanut, vetches and pigeonpea, its sequence is important for the study of the function of genes within this economically important group of legumes. In relation to current development of sequencing methodology, there is issue if to use whole genome shotgun (WGS) method based on the Sanger technique, or BAC clone approach. New approaches such as the 454 Roche pyrosequencing or Illumina, offer a cost advantage along with increased speed and throughput. Considering the large proportion of repetitive sequences and size of pea genome, it will be important to have sufficient genetic and physical tools for scaffold assembly and merging the scaffolds into pseudochromosomes. There is community-wide effort with input and support from many individuals resulting in establishment of an International Consortium for Pea Genome Sequencing (http://www.gabcsfl.org). Scientists and breeders might profit from knowledge of pea genome similarly to pigeonpea, recently sequenced by Illumina platform to generate 237.2 Gb of sequence, along with Sanger-based bacterial artificial chromosome end sequences and a genetic map, assembled into scaffolds representing 72.7 % (605.78 Mb) of the 833.07 Mb pigeonpea genome (Varhney et al. 2012; Schmutz et al. 2010).

Conclusion

Knowledge of pea genome architecture will facilitate the identification of a wide range of DNA markers, genes, and pea genotypes that influence important traits such as resistance to biotic and abiotic stress; yield stability and nutritional quality. Newly identified genes and alleles controlling these traits will enable marker-assisted breeding and transgenic strategies for accelerating pea enhancement. Importantly, genomic knowledge would allow application of linkage disequilibrium (LD) or association mapping strategy. As shown above the extension of model legumes for comparative functional genomics, together with “omics” knowledge, is starting to provide candidate genes for QTL identification of genes involved in stress and quality traits. As genes are identified in model legumes and crop species comparison and transfer of candidate gene information from the model to the crop species is possible, favourable alleles for breeding and selection will be identified, and improved varieties will be developed by marker assisted selection (MAS) or genetic transformation.

References

Abbo S, Lev Yadun S, Gopher A (2010) Agricultural origins: centers and noncenters; a near eastern reappraisal. Crit Rev Plant Sci 29:317–328

Ambrose MJ (1995) From near east centre of origin the prized pea migrates throughout world. Diversity 11:118–119

Andrade M, Abe Y, Nakahara KS, Uyeda I (2009) The *cyv-2* resistance to *Clover yellow vein virus* in pea is controlled by the eukaryotic initiation factor 4E. J Gen Plant Pathol 75:241–249

Atif RM, Patat-Ochatt EM, Svabova L, Ondrej V, Klenoticova H, Jacas L, Griga M, Ochatt S (2013) Gene transfer in legumes. Luttge U et al (eds) Progress in botany, vol 74. Springer, Berlin

Attanayake RN, Glawe DA, McPhee KE et al (2010) *Erysiphe trifolii* – a newly recognized powdery mildew pathogen of pea. Plant Pathol 59:712–720

Aubert G, Morin J, Jacquin F, Loridon K, Quillet MC, Petit A, Rameau C, Lejeune Hénaut I, Huguet T, Burstin J (2006) Functional mapping in pea, as an aid to the candidate gene selection and for investigating synteny with the model legume *Medicago truncatula*. Theor Appl Genet 112:1024–1041

Baranger AG, Aubert G, Arnau G, Lainé AL, Deniot G, Potier J, Weinachter C, Lejeune-Hénaut J, Lallemand J, Burstin J (2004) Genetic diversity within *Pisum sativum* using protein- and PCR-based markers. Theor Appl Genet 108:1309–1321

Barba-Espín G, Diaz-Vivancos P, Job D, Belghazi M, Job C, Hernández JA (2011) Understanding the role of H_2O_2 during pea seed germination: a combined proteomics and hormone profiling approach. Plant Cell Environ 34:1907–1919

Bardel J, Louwagie M, Jaquinod M, Jourdain A, Luche S, Rabilloud T, Macherel D, Garin J, Bourguignon J (2002) A survey of the plant mitochondrial proteome in relation to development. Proteomics 2:880–898

Barilli E, Satovic Z, Rubiales D, Torres AM (2010) Mapping of quantitative trait loci controlling partial resistance against rust incited by *Uromyces pisi* (Pers.) Wint. in a *Pisum fulvum* L. intraspecific cross. Euphytica 175:151–159

Barilli E, Rubiales D, Castillejo A (2012) Comparative proteomic analysis of BRTH and BABA-induced resistance in pea (*Pisum sativum*) toward infection in pea rust (*Uromyces pisi*). J Proteomics 75:5189–5205

Bastianelli D, Grosjean F, Peyronnet C, Duparque M, Regnier JM (1998) Feeding value of pea (*Pisum sativum* L.) – 1 Chemical composition of different categories of pea. Animal Sci 67:609–619

Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. Nat Rev Genet 5:251–261

Blixt S (1972) Mutation genetics in *Pisum*. Agric Hort Genet 30:1–293

Bordat A, Savois V, Nicolas M, Salse J, Chauveau A, Bourgeois M, Potier J, Houtin H, Rond C, Murat F, Marget P, Aubert G, Burstin J (2011) Translational genomics in legumes allowed placing *insilico* 5460 unigenes on the pea functional map and identified candidate genes in *Pisum sativum* L. G3 2:93–103

Bourgeois M, Jacquin F, Cassecuelle F, Savois V, Belghazi M, Aubert G, Quillien L, Huart M, Marget P, Burstin J (2011a) A PQL (protein quantity loci) analysis of mature pea seed proteins identifies loci determining seed protein. Proteomics 11:1581–1594

Bourgeois M, Jacquin F, Cassecuelle F, Savois V, Belghazi M, Aubert G, Quillien L, Huart M, Marget P, Burstin J (2011b) A PQL (protein quantity loci) analysis of mature pea seed proteins identifies loci determining seed protein composition. Proteomics 9:1581–1594

Bourgeois M, Jacquin F, Savois V, Sommerer N, Labas V, Henry C, Burstin J (2009) Dissecting the proteome of pea mature seeds reveals the phenotypic plasticity of seed protein composition. Proteomics 9:254–271

Bourion V, Rizvi SMH, Fournier S, de Larambergue H, Galmiche F, Marget P, et al (2010) Genetic dissection of nitrogen nutrition in pea through a QTL approach of root, nodule, and shoot variability. Theor Appl Genet 121:71–86

Bruun-Rasmussen M, Møller IS, Tulinius G, Hansen KR, Lund OS, Johansen IE (2007) The same allele of translation initiation factor 4E mediates resistance against two *Potyvirus* spp. in *Pisum sativum*. Mol Plant Microbe Interact 20:1075–1082

Burstin J, Deniot G, Potier J, Weinachter C, Aubert G, Baranger A (2001) Microsatellite polymorphism in *Pisum sativum*. Plant Breed 120:311–317

Burstin J, Marget P, Huart M, Moessner A, Mangin B, Duchene C, Desprez B, Munier-Jolain N, Duc G (2007) Developmental genes have pleiotropic effects on plant morphology and source capacity, eventually impacting on seed protein content and productivity in pea. Plant Physiol 144:768–781

Byrne OM, Hardie DC, Khan T, Yan G (2008) Genetic analysis of pod and seed resistance to pea weevil in a *Pisum sativum* x *P. fulvum* interspecific cross. Aust J Agric Res 59:854–862

Cannon SB, May GD, Jackson SA (2009) Three sequenced legume genomes and many crop species: rich opportunities for translational genomics. Plant Physiol 151:970–977

Castillejo MA, Amiour N, Gaudot ED, Rubiales D, Jorrín JV (2004) A proteomic approach to studying plant response to crenate broomrape (*Orobanche crenata*) in pea (*Pisum sativum*). Phytochemistry 65:1817–1828

Charlton A, Allnutt T, Hølmes S, Chisholm J, Bean S, Ellis N, Mullineaux P, Oehlschlager S (2004) NMR profiling of transgenic peas. Plant Biotech J 2:27–35

Charlton A, Donarski JA, Harrison M, Jones SA, Godward J, Oehlschlager S, Arques JL, Ambrose M, Chinoy C, Mullineaux PM, Domoney C (2008) Responses of the pea (*Pisum sativum* L.) leaf metabolome to drought stress assessed by nuclear magnetic resonance spectroscopy. Metabolomics 4:312–327

Chinoy C, Welham T, Turner L, Moreau C, Domoney C (2011) The genetic control of seed quality traits: effects of allelic variation at the *Tri* and *Vc-2* genetic loci in *Pisum sativum* L. Euphytica 180:107–122

Choi HK, Mun JH, Kim DJ, Zhu HY, Baek JM, Mudge J, Bruce R, Noel E, Jeff D, Gyorgy KB, Young ND, Cook Douglas R (2004) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci U S A 101:15289–15294

Chowrira GM, Cavileer TD, Gupta SK, Lurquin PF, Berger PH (1998) Coat protein mediated resistance to pea enation mosaic virus in transgenic *Pisum sativum* L. Transgenic Res 7:265–271

Cieslarová J, Hanáček P, Fialová E, Hýbl M, Smýkal P (2011) Estimation of pea (*Pisum sativum* L.) microsatellite mutation rate based on pedigree and single-seed descent analyses. J Appl Genet 52:391–401

Clement SL, McPhee KE, Elberson LR, Evans MA (2009) Pea weevil, *Bruchus pisorum* L. (Coleoptera: Bruchidae), resistance in *Pisum sativum*×*Pisum fulvum* interspecific crosses. Plant Breed 12:478–486

Clulow SA, Matthews P, Lewis BG (1991) Genetical analysis of resistance to *Mycosphaerella pinodes* in pea seedlings. Euphytica 58:183–189

Constantin GD, Krath BN, MacFarlane SA, Nicolaisen M, Johansen IE, Lund OS (2004) Virus-induced gene silencing as a tool for functional genomics in a legume species. Plant J 40: 622–631

Corander J, Waldmann P, Sillanpää MJ (2003) Bayesian analysis of genetic differentiation between populations. Genetics 164:367–374

Coyne CJ, Inglis DA, Whitehead SJ, Muehlbauer FJ (2000) Chromosomal location of *Fwf* in pea. Pisum Genet 32:20–22

Coyne CJ, Mc Clendon MT, Walling JG, Timmerman-Vaughan GM, Murray S, Meksem K, Lightfoot DA, Shultz JL, Keller KE, Martin RR, Inglis DA, Rajesh PN, McPhee KE, Weeden NF, Grusak MA, Li CM, Storlie EW (2007a) Construction and characterization of two bacterial artificial chromosome libraries of pea (*Pisum sativum* L.) for the isolation of economically important genes. Genome 50:871–875

Coyne CJ, Pilet-Nayel ML, Grünwald NJ (2007) Identification of QTLs controlling resistance to Fusarium root rot in pea. 6th European conference on grain legumes from 12 to 16 Nov 2007, Lisbon, Portugal

Darby P, Lewis BG, Matthews P (1985) Inheritance and expression of resistance to *Ascochyta pisi*. In: Hebblethwaite PD, Heath MC, Dawkins TC, (eds). The pea crop. London: Butterworths p. 231–236

Dalmais M, Schmidt J, Le Signor C, Moussy F, Burstin J, Savois V, Aubert G, Brunaud V, de Oliveira Z, Guichard C, Thompson R, Bendahmane A (2008) UTILLdb, a *Pisum sativum* in silico forward and reverse genetics tool. Genome Biol 9:R43

De Candolle A (1882) Origin of cultivated plants. Kesinger Publishing LCC, Whitefish, MT

Decarie J, Coyne C, Brumett S, Shultz J (2012) Additional pea EST-SSR markers for comparative mapping in pea (*Pisum sativum* L.). Plant Breed 131:222–226

DeMason DA, Chetty VJ (2011) Interaction between GA, auxin, and UNI expression controlling shoot ontogeny, leaf morphogenesis, and auxin response in *Pisum sativum* (*Fabaceae*): or how the uni-tac mutant is rescued. Am J Bot 98:775–791

Deulvot C, Charrel H, Marty A, Jacquin F, Donnadieu C, Lejeune-Hénaut I, Burstin J, Aubert G (2010) Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. BMC Genomics 11:468

Dirlewanger E, Isaac P, Ranade S, Belajouza M, Cousin R, Devienne D (1994) Restriction fragment length polymorphism analysis of loci associated with disease resistance genes and developmental traits in *Pisum sativum* (L.). Theor Appl Genet 88:17–27

Dixon RA, Sumner LW (2003) Legume natural products: understanding and manipulating complex pathways for human and animal health. Plant Physiol 131:878–885

Dolezel J, Greilhuber J (2010) Nuclear genome size. Are we getting closer? Cytometry 77:635–642

Doss RP, Oliver JE, Proebsting WM, Potter SW, Kuy SR, Clement SL, Williamson RT, Carney JR, De Vilbiss ED (2000) Bruchins: insect-derived plant regulators that stimulate neoplasm formation. Proc Natl Acad Sci U S A 97:6218–6223

Doss RP, Proebsting WM, Potter SW, Clement SL (1995) Response of *np* mutant of pea (*Pisum sativum* L.) to pea weevil (*Bruchus pisorum* L) oviposition and extracts. J Chem Ecol 21:97–106

Dumont E, Fontaine V, Vuylsteker C, Sellier H, Bodèle S, Voedts N, Devaux R, Frise M, Avia K, Hilbert JL, Bahrman N, Hanocq E, Lejeune-Hénaut I, Delbreil B (2009) Association of sugar content QTL and PQL with physiological traits relevant to frost damage resistance in pea under field and controlled conditions. Theor Appl Genet 118:1561–1571

Ek M, Eklund M, Von Post R, Dayteg C, Henriksson T, Weibull P, Ceptilis A, Issac P, Tuvesson S (2005) Microsatellite markers for powdery mildew resistance in pea (*Pisum sativum* L.). Hereditas 142:86–91

Ellis THN, Poyser SJ, Knox MR, Vershinin AV, Ambrose MJ (1998) Polymorphism of insertion sites of Ty1-*copia* class retro-transposons and its use for linkage and diversity analysis in pea. Mol Gen Genet 260:9–19

Ellis THN, Poyser SJ (2002) An integrated and comparative view of pea genetic and cytogenetic maps. New Phytologist 153:17–25

Ellis THN (2011) *Pisum*. In: Kole C (ed) Wild crop relatives: genomics and breeding resources, Chapter 12. Springer, Berlin, pp 237–248

Elvira-Recuenco M, Taylor JD (2001) Resistance to bacterial blight (*Pseudomonas syringae* pv. *pisi*) in Spanish pea (*Pisum sativum*) landraces. Euphytica 118:305–311

Fann YJ, Li W, Wang JJ, Liu JY, Yang MY, Xu D, Yhu XJ, Wang XZ (2011) Efficient production of human acidic fibroblast growth factor in pea (*Pisum sativum* L.) plants by agroinfection of germinated seeds. BMC Biotechnol 11:45

FAOSTAT. Food and Agriculture Organization of the United Nations (2012). http://faostat.fao.org/site/567/default.aspx#ancor. Accessed 10 Jan 2013

Flavell AJ, Bolshakov VN, Booth A, Jing R, Russell J, Ellis THN, Isaac P (2003) A microarray-based high throughput molecular marker genotyping method: the tagged microarray marker (TAM) approach. Nucl Acids Res 31:e115

Fondevilla S, Almeida NF, Satovic Z, Rubiales D, Vaz Patto MC, Cubero JI, Torres AM (2011a) Identification of common genomic regions controlling resistance to *Mycosphaerella pinodes*, earliness and architectural traits in different pea genetic backgrounds. Euphytica 18:43–52

Fondevilla S, Carver TLW, Moreno MT, Rubiales D (2006) Macroscopic and histological characterisation of genes *er1* and *er2* for powdery mildew resistance in pea. Eur J Plant Pathol 115:309–321

Fondevilla S, Küster H, Krajinski F, Cubero JI, Rubiales D (2011a) Identification of genes differentially expressed in a resistant reaction to *Mycosphaerella pinodes* in pea using microarray technology. BMC Genomics 12:28

Fondevilla S, Küster H, Krajinski F, Cubero JI, Rubiales D (2011b) Identification of genes differentially expressed in a resistant reaction to *Mycosphaerella pinodes* in pea using microarray technology. BMC Genomics 12:28

Fondevilla S, Martın-Sanz A, Satovic Z, Fernández-Romero MD, Rubiales D, Caminero C (2012) Identification of quantitative trait loci involved in resistance to *Pseudomonas syringae* pv. *syringae* in pea (*Pisum sativum* L.). Euphytica 186:805–812

Fondevilla S, Rubiales D, Moreno MT, Torres AM (2008) Identification and validation of RAPD and SCAR markers linked to the gene *Er3* conferring resistance to *Erysiphe pisi* DC in pea. Mol Breed 22:193–200

Fondevilla S, Torres AM, Moreno MT, Rubiales D (2007) Identification of a new gene for resistance to powdery mildew in *Pisum fulvum*, a wild relative of pea. Breed Sci 57:181–184

Ford R, Le Roux K, Itman C, Brouwer JB, Taylor PWJ (2002) Diversity analysis and genotyping in *Pisum* with sequence tagged microsatellite site (STMS) markers. Euphytica 124:397–405

Foucher F, Morin J, Courtiade J, Cadioux S, Ellis N, Banfield MJ, Rameau C (2003) Determinate and late flowering are two terminal flower centroradialis homologs that control two distinct phases of flowering initiation and development in pea. Plant Cell 15:2742–2754

Franssen SU, Shrestha RP, Bräutigam A, Bornberg-Bauer E, Weber APM (2011) Comprehensive transcriptome analysis of the highly complex *Pisum sativum* genome using next generation sequencing. BMC Genomics 12:227

Gao Z, Eyers S, Thomas C, Ellis N, Maule A (2004) Identification of markers tightly linked to *sbm* recessive genes for resistance to *Pea seed-borne mosaic virus*. Theor Appl Genet 109:488–494

Ghafoor A, McPhee K (2012) Marker assisted selection (MAS) for developing powdery mildew resistant pea cultivars. Euphytica 186:593–607

Gilpin BJ, McCallum JA, Frew TJ, Timmerman-Vaughan GM (1997) A linkage map of the pea (*Pisum sativum* L.) genome containing cloned sequences of known function and expressed sequence tags (ESTs). Theor Appl Genet 95:1289–1299

Gong Y, Xu S, Mao W, Hu Q, Zhang G, Ding J, Li Y (2010) Developing new SSR markers from ESTs of pea (*Pisum sativum* L.). J Zhejiang Univ-Sci B 11:702–707

Grajal-Martin MJ, Muehlbauer FJ (2002) Genomic location of the *Fw* gene for resistance to fusarium wilt race 1 in peas. J Heredity 93:291–293

Grelet J, Benamar A, Teyssier E, Avelange-Macherel MH, Grunwald D, Macherel D (2005) Identification in pea seed mitochondria of late-embryogenesis abundant protein able to protect enzymes from drying. Plant Physiol 137:157–167

Gronlund M, Olsen A, Johansen IE, Jakobsen I (2010) Protocol: using virus-induced gene silencing to study the arbuscular mycorrhizal symbiosis in *Pisum sativum*. Plant Methods 6:28

Guo SM, Kamphuis LG, Gao LL, Klingler JP, Lichtenzveig J, Edwards O, Singh KB (2012) Identification of distinct quantitative trait loci associated with defence against the closely related aphids *Acyrthosiphon pisum* and *A. kondoi* in *Medicago truncatula*. J Exp Bot 63:3913–3922

Hall KJ, Parker JS, Ellis TH (1997) The relationship between genetic and cytogenetic maps of pea. I. Standard and translocation karyotypes. Genome 40:744–754

Hamon C, Baranger A, Coyne CJ, McGee RJ, Le Goff I, L'Anthoëne V, Esnault R, Rivière JP, Klein A, Mangin P, McPhee KE, Roux-Duparque M, Porter L, Miteul H, Lesné A, Morin G, Onfroy C, Moussart A, Tivoli B, Delourme R, Pilet-Nayel ML (2010) New consistent QTL in pea associated with partial resistance to *Aphanomyces euteiches* in multiple field and controlled environments from France and the United States. Theor Appl Genet 123:261–281

Hardie DC, Baker GJ, Marshall DR (1995) Field screening of *Pisum* accessions to evaluate their susceptibility to the pea weevil (Coleoptera: Bruchidae). Euphytica 84:155–161

Harland SC (1948) Inheritance of immunity to mildew in Peruvian forms of *Pisum sativum*. Heredity 2:263–269

Hecht V, Knowles CL, Vander Schoor JK, Liew LC, Jones SE, Lambert MJ, Weller JL (2007) Pea LATE BLOOMER1 is a GIGANTEA ortholog with roles in photoperiodic flowering, deetiolation, and transcriptional regulation of circadian clock gene homologs. Plant Physiol 144:648–661

Hecht V, Laurie RE, Vander Schoor JK, Ridge S, Knowles CL, Liew LC, Sussmilch FC, Murfet IC, Macknight RC, Weller JL (2011) The pea GIGAS gene is a FLOWERING LOCUS T homolog necessary for graft-transmissible specification of flowering but not for responsiveness to photoperiod. Plant Cell 23:147–161

Hellens RP, Moreau C, Lin-Wang K, Schwinn KE, Thomson SJ, Fiers MWEJ, Frew TJ, Murray SR, Hofer JMI, Jacobs JME, Davies KM, Allan AC, Bendahmane A, Coyne CJ, Timmerman-Vaughan GM, Ellis THN (2010) Identification of Mendel's white flower character. PLoS ONE 5:e13230

Heringa RJ, Van Norel A, Tazelaar MF (1969) Resistance to powdery mildew (*Erysiphe polygoni* D.C.) in peas (*Pisum sativum* L.). Euphytica 18:163–169

Humphry M, Reinstaedler A, Ivanov S, Bisseling T, Panstruga R (2011) Durable broad-spectrum powdery mildew resistance in pea *er1* plants is conferred by natural loss-of-function mutations in PsMLO1. Mol Plant Pathol 12:866–878

Hunter PJ, Ellis N, Taylor JD (2001) Association of dominant loci for resistance to *Pseudomonas syringae* pv. *pisi* with linkage groups II, VI and VII of *Pisum sativum*. Theor Appl Genet 103:129–135

Irzykowska L, Wolko B, Swiêcicki WK (2001) The genetic linkage map of pea (*Pisum sativum* L.) based on molecular, biochemical and morphological markers. Pisum Genet 33:13–18

Irzykowska L, Wolko B (2004) Interval mapping of QTLs controlling yield-related traits and seed protein content in *Pisum sativum*. J Appl Genet 45:297–306

Jing R, Ambrose MA, Knox MR, Smýkal P, Hybl M, Ramos A et al (2012) Genetic diversity in European Pisum germplasm collections. Theor Appl Genet 125:367–380

Jing R, Johnson R, Seres A, Kiss G, Ambrose MJ, Knox MR, Ellis THN, Flavell AJ (2007) Gene-based sequence diversity analysis of field pea (*Pisum*). Genetics 177:2263–2275

Jing R, Vershinin A, Grzebyta J, Shaw P, Smýkal P, Marshall D, Ambrose MJ, Ellis THN, Flavell AJ (2010) The genetic diversity and evolution of field pea (*Pisum*) studied by high throughput retro-transposon based insertion polymorphism (RBIP) marker analysis. BMC Evol Biol 10:44

Jing RC, Knox MR, Lee JM, Vershinin AV, Ambrose M, Ellis THN, Flavell AJ (2005) Insertional polymorphism and antiquity of *PDR1* retrotransposon insertions in *Pisum* species. Genetics 171:741–752

Johansen EI, Lund OS, Hjulsager CK, Laursen J (2001) Recessive resistance in *Pisum sativum* and Potyvirus pathotype resolved in a gene-for-cistron correspondence between host and virus. J Virol 75:6609–6614

Kaimoyo E, Van Etten HD (2008) Inactivation of pea genes by RNAi supports the involvement of two similar O-methyltransferases in the biosynthesis of (+)-pisatin and of chiral intermediates with a configuration opposite that found in (+)-pisatin. Phytochemistry 69:76–87

Kalo P, Seres A, Taylor SA, Jaka J, Kevei ZA, Kereszt GE, Ellis THN, Kiss GB (2004) Comparative mapping between *Medicago sativa* and *Pisum sativum*. Mol Gen Genom 272:235–246

Kanervo E, Singh M, Suorsa M, Paakkarinen V, Aro E, Battchikova N, Aro EM (2008) Expression of protein complexes and individual proteins upon transitiv of etioplasts to chloroplasts in pea (*Pisum sativum*). Plant Cell Physiol 49:396–410

Katoch V, Sharma S, Pathania S, Banayal K, Sharma SK, Rathour R (2010) Molecular mapping of pea powdery mildew resistance gene *er2* to pea linkage group III. Mol Breed 25:229–237

Kaur S, Pembleton LW, Cogan NO, Savin KW, Leonoforte T, Paull J, Materne M, Forste JW (2012) Transcriptome sequencing of field pea and fababean for discovery and validation of SSR genetic markers. BMC Genomics 13:104

Krajewski P, Bocianowski J, Gawłowska M, Kaczmarek Z, Pniewski T, Swiecicki W, et al (2012) QTL for yield components and protein content: a multienvironment study of two pea (*Pisum sativum* L.) populations. Euphytica 183:323–336

Kwon SJ, Brown AF, Hu J, McGee RJ, Watt CA, Kisha T, et al. (2012) Population genetic sub-structure within the USDA ARS *Pisum* core collection and its potential as a platform for association mapping. Genes Genomics 34:305–320

Kwon SJ, Smykal P, Hu J, McGee RJ, McPhee KE, Coyne CJ (2013) User-friendly markers linked to *Fusarium* wilt race 1 resistance gene in pea for marker-assisted selection. Plant Breed doi:10.111/pbr.12085

Knox MR, Ellis THN (2002) Excess heterozygosity contributes to genetic map expansion in pea recombinant inbred populations. Genetics 162:861–873

Knox MR, Ellis THN (2001) Stability and inheritance of methylation states at *PstI* sites in *Pisum*. Mol Genet Genomics 265:497–507

Konovalov F, Toshchakova E, Gostimsky S (2005) A CAPS marker set for mapping in linkage group III of pea (*Pisum sativum* L.). Cell Mol Biol Lett 10:163–171

Kraft JM, Pfleger FL (2001) Compendium of pea diseases and pests. APS Press, St. Paul, MN, pp 14–17

Krusell L, Sato N, Fukuhara I, Koch EBV, Grossmann C, Okamoto S, Oka-Kira E, Otsubo Y, Aubert G, Nakagawa T, Sato S, Tabata S, Duc G, Parniske M, Wang TL, Kawaguchi M, Stougaard J (2008) The Clavata2 genes of pea and *Lotus japonicus* affect autoregulation of nodulation. Plant J 65:861–871

Künne C, Lange M, Funke T, Miehe H, Thiel T, Grosse I, Scholz U (2005) CR-EST: a resource for crop ESTs. Nucleic Acids Res 33:619–621

Kushwaha C, Chand R, Srivastava C (2006) Role of aeciospores in outbreaks of pea (*Pisum sativum*) rust (*Uromyces fabae*). Eur J Plant Path 115:323–330

Laucou V, Haurogne K, Ellis N, Rameau C (1998) Genetic mapping in pea. I. RAPD-based linkage map of *Pisum sativum*. Theor Appl Genet 97:905–915

Lejeune-Hénaut I, Hanocq E, Béthencourt L, Fontaine V, Delbreil B, Morin J, Petit A, Devaux R, Boilleau M, Stempniak JJ, Thomas M, Lainé AL, Foucher F, Baranger A, Burstin J, Rameau C, Giauffret C (2008) The flowering locus Hr colocalizes with a major QTL affecting winter frost tolerance in *Pisum sativum* L. Theor Appl Genet 116:1105–1116

Liang D, Wong CE, Singh MB, Beveridge CA, Phipson B, Smyth GK, Bhalla PL (2009) Molecular dissection of the pea shoot apical meristem. Exp Bot 60:4201–4213

Liew L (2011) Genetic control of photoperiod responsiveness in pea. Ph.D. thesis, University of Tasmania, Australia

Loridon K, McPhee KE, Morin J, Dubreuil P, Pilet-Nayel ML, Aubert G, Rameau C, Baranger A, Coyne CJ, Lejeune-Hénault I, Burstin J (2005) Microsatellite marker polymorphism and mapping in pea (*Pisum sativum* L.). Theor Appl Genet 111:1022–1031

Lucau-Danilla A, Toitot C, Goulas E, Blervacq AS, Hot D, Bahrman N, Sellier H, Lejeune-Hénaut I, Delbreil B (2012) Transcriptome analysis in pea allows distinguishing chilling and acclimation mechanisms. Plant Physiol Biochem 58:236–244

Macas J, Neumann P, Navrátilová A (2007) Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. BMC Genomics 8:427

Matamoros MA, Loscos J, Dietz K-J, Aparicio-Tejo PM, Becana M (2009) Function of antioxidant enzymes and metabolites during maturation of pea fruits. J Exp Bot 61:87–97

Maxted N, Ambrose N (2000) Plant genetic resources of legumes in the Mediterranean, Chapter 10. In: Maxted N, Bennett SJ (eds) Peas (*Pisum* L.). Kluwer Academic Publishers, Dordrecht, pp 181–190

McCallum J, Timmerman-Vaughan G, Frew T, Russell A (1997) Biochemical and genetic linkage analysis of green seed color in field pea. J Am Hort Sci 122:218–225

McClendon MT, Inglis DA, McPhee KE, Coyne CJ (2002) DNA markers for *Fusarium* wilt race 1 resistance gene in pea. J Am Soc Hort Sci 127:602–607

McPhee K (2007) Pea, Chapter 2. In: Kole C (ed) Pulses, sugar and tuber crops. Genome mapping and molecular breeding in plants, vol 3. Springer, Berlin, pp 33–47

McPhee KE, Inglis DA, Gunderson B, Coyne CJ (2012) Mapping a resistance gene for *Fusarium* wilt Race 2 on LG IV of pea (*Pisum sativum* L.). Plant Breed, doi:10.1111/j.1439-0523.2011.01938.x

Mikić A, Mihailović V, Ćupina B, Kosev V, Warkentin T, McPhee K, Ambrose M, Hofer J, Ellis N (2011) Genetic background and agronomic value of leaf types in pea (*Pisum sativum*). Field Veg Crops Res 48:275–284

Mikschofsky H, Schirrmeier H, Keil GM, Lange B, Polowick P, Keller W, Broer I (2009) Pea derived vaccines demonstrate high immunogenicity and protection in rabbits against haemorrhagic disease virus. Plant Biotechnol 7:537–549

Mishra RK, Gangadhar BH, Nookaraju A, Kumar S, Park SW (2012) Development of EST-derived SSR markers in pea (*Pisum sativum*) and their potential utility for genetic mapping and transferability. Plant Breed 131:118–124

Molnár Z, Jenes B, Ordog VV (1999) Genetic transformation of pea (*Pisum sativum* L.) via particle bombardment. 3rd International symposium in the series recent advances in plant biotechnology. September 4–10, Stara Lesna, Slovakia. Biologia 54:50

Morton RL, Schroeder HE, Bateman K et al (2000) Bean a-amylase inhibitor 1 in transgenic peas (*Pisum sativum* L.) provides complete protection from pea weevil (*Bruchus pisorum*) under field conditions. Proc Natl Acad Sci U S A 7:3820–3825

Murfet IC (1973) Flowering in *Pisum*. *Hr*, a gene for high response to photoperiod. Heredity 31:157–164

Murfet IC (1975) Flowering in *Pisum*. Multiple alleles at the Lf locus. Heredity 35:85–98

NCBI. National center for biotechnology information. http://www.ncbi.nlm.nih.gov. Accessed 29 Jan, 2013

Neumann P, Nouzová M, Macas J (2001) Molecular and cytogenetic analysis of repetitive DNA in pea (*Pisum sativum* L.). Genome 44:716–728

Okubara PA, Keller KE, McClendon MT, McPhee KE, Inglis DA, Coyne CJ (2005) Y15_999Fw, a dominant SCAR marker linked to the Fusarium wilt race 1 (*Fw*) resistance gene in pea. Pisum Genet 37:32–35

Ondřej M, Dostálová R, Odstrčilová L (2005) Response of *Pisum sativum* germplasm resistant to *Erysiphe pisi* to inoculation with *Erysiphe baeumleri*, a new pathogen of pea. Plant Protect Sci 41:95–103

Page D, Aubert G, Duc G, Welham T, Domoney C (2002) Combinatorial variation in coding and promoter sequences of genes at the *Tri* locus in *Pisum sativum* accounts for variation in trypsin inhibitor activity in seeds. Mol Genet Genom 267:359–369

Pavan S, Schiavulli A, Appiano M, Marcotrigiano AR, Cillo F, Visser RGF, Bai Z, Lotti C, Ricciardi L (2011) Pea powdery mildew *er1* resistance is associated to loss-of-function mutations at a *MLO* homologous locus. Theor Appl Genet 123:1425–1431

Pereira G, Marques C, Ribeiro R, Formiga S, Damaso M, Tavares-Sousa M, Farinho M, Leitão JM (2010) Identification of DNA markers linked to an induced mutated gene conferring resistance to powdery mildew in pea (*Pisum sativum* L.). Euphytica 171:327–335

Perrin Y, Vaquero C, Gerrad I, Sack M, Drossard J, Stoger E, Christou P, Fisher R (2000) Transgenic pea seeds as bioreactors for production of a single chain Fv fragment (scFV) antibody used in cancer diagnosis and therapy. Mol Breed 6:345–352

Pilet-Nayel ML, Muehlbauer FJ, McGee RJ, Kraft JM, Baranger A, Coyne CJ (2002) Quantitative trait loci for partial resistance to *Aphanomyces* root rot in pea. Theor Appl Genet 106:28–39

Pilet-Nayel ML, Muehlbauer FJ, McGee RJ, Kraft JM, Baranger A, Coyne CJ (2005) Consistent QTL in pea for partial resistance to *Aphanomyces euteiches* isolates from United States and France. Phytopathol 95:1287–1293

Pradhan S, Adams RLP (1995) Distinct CG and CNG methyltransferases in *Pisum sativum*. Plant J 7:471–481

Prescott VE, Campbell PM, Moore A, Mattes J, Rothenberg ME, Foster PS, Higgins TJV, Hogan SP (2005) Transgenic expression of bean alpha-amylase inhibitor in peas results in altered structure and immunogenicity. J Agric Food Chem 53:9023–9030

Prioul S, Frankewitz A, Deniot G, Morin G, Baranger A (2004) Mapping of quantitative trait loci for partial resistance to *Mycosphaerella pinodes* in pea (*Pisum sativum* L.), at the seedling and adult plant stages. Theor Appl Genet 108:1322–1334

Provvidenti R, Hampton RO (1991) Chromosomal distribution of genes for resistance to seven potyviruses in *Pisum sativum*. Pisum Genet 23:26–28

Puonti-Kaerlas J, Eriksson T, Engstroem P (1999) Production of transgenic pea (*Pisum sativum* L.) plants by Agrobacterium-mediated gene transfer. Theor Appl Genet 80:246–252

Radchuk R, Emery NJR, Weier D, Vigeolas H, Geigenberger P, Lunn JE, Feil R, Weschke W, Weber H (2010) Sucrose non-fermenting kinase 1 (*SnRK1*) coordinates metabolic and hormonal signals during pea cotyledon growth and differentiation. Plant J 61:324–338

Rai R, Singh AK, Singh BD, Joshi AK, Chand R, Srivastava CP (2011) Molecular mapping for resistance to pea rust caused by *Uromyces fabae* (Pers.) de-Bary. Theor Appl Genet 123:803–813

Ramachandran VK, East AK, Karunakaran R, Downie JA, Poole PS (2011) Adaptation of *Rhizobium leguminosarum* to pea, alfalfa and sugar beet rhizosphere investigated by comparative transcriptomics. Genome Biol 12:106

Reif JC, Melchinger AE, Frisch M (2005) Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. Crop Sci 45:1–7

Riebeseel E, Häusler RE, Radchuk R, Meitzel T, Hajirezaei MR, Emery RJN, Küster H, Nunes-Nesi A, Fernie AR, Weschke W, Weber H (2010) The 2-oxoglutarate/malate translocator mediates amino acid and storage protein biosynthesis in pea embryos. Plant J 61:350–363

Rubiales D, Fernández-Aparicio M, Pérez-de-Luque A, Prats E, Castillejo MA, Sillero J, Rispail N, Fondevilla S (2009) Breeding approaches for crenate broomrape (*Orobanche crenata* Forsk.) management in pea (*Pisum sativum* L.). Pest Manag Sci 65:553–559

Rubiales D, Moreno MT, Sillero JC (2005) Search for resistance to crenate broomrape (*Orobanche crenata*) in pea germplasm. Genet Res Crop Evol 52:853–861

Saalbach I, Giersberg M, Conrad U (2001) High-level expression of a single-chain Fv fragment (scFV) antibody in transgenic pea seeds. J Plant Physiol 158:529–533

Saha GC, Vandemark GJ (2012) Evaluation of expression stability of candidate references genes among green and yellow pea cultivars (*Pisum sativum* L.) subjected to abiotic and biotic stress. Am J Plant Sci 3:235–242

Schiltz S, Gallardo K, Huart M, Negroni L, Sommerer N, Burstin J (2004) Proteome reference maps of vegetative tissues in pea. An investigation of nitrogen mobilization from leaves during seed filling. Plant Physiol 135:2241–2260

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Schroeder HE, Gollash S, Tabe LM, Higgins TJV (1994) Recent advances in gene transfer to peas. Pisum Genet 26:1–5

Schroeder HE, Scholtz AH, Wardley-Richardson T, Spencer D, Higgins TJV (1993) Tranformation and regeneration of two cultivars of pea (*Pisum sativum* L.). Plant Physiol 101:751–757

Simon CJ, Muehlbauer FJ (1997) Construction of a chickpea linkage map and its comparison with maps of pea and lentil. J Heredity 88:115–119

Smartt J (2005) Grain legumes: evolution and genetic resources. Cambridge University Press, Cambridge

Smýkal P, Aubert G, Burstin J, Coyne C, Ellis N, Flavell A, Ford R, Hýbl M, Macas J, Neumann P, McPhee K, Redden R, Rubiales D, Weller J, Warkentin TD (2012a) Pea (*Pisum sativum* L.) in the genomic era. Agronomy 2:74–115

Smýkal P, Coyne CJ, Ford R, Redden R, Flavell AJ, Hýbl M, Warkentin T, Burstin J, Duc G, Ambrose M, Ellis THN (2008a) Effort towards a world pea (*Pisum sativum* L.) germplasm core collection: the case for common markers and data compatibility. Pisum Genet 40:11–14

Smýkal P, Hýbl M, Corander J, Jarkovský J, Flavell AJ, Griga M (2008b) Genetic diversity and population structure of pea (*Pisum sativum* L.) varieties derived from combined retro-transposon, microsatellite and morphological marker analysis. Theor Appl Genet 117:413–424

Smýkal P, Kalendar R, Ford R, Macas J, Griga M (2009) Evolutionary conserved lineage of *Angela*-family retrotransposons as a genome-wide microsatellite repeat dispersal agent. Heredity 103:157–167

Smýkal P, Kenicer G, Flavell AJ, Corander J, Kosterin O, Redden RJ, Ford R, Coyne CJ, Maxted N, Ambrose MJ, Ellis THN (2011) Phylogeny, phylogeography and genetic diversity of the *Pisum* genus. Plant Genet Res 9:4–18

Smýkal P, Šafářová D, Navrátil M, Dostalová R (2010) Marker assisted pea breeding: *eIF4E* allele specific markers to *Pea seed-borne mosaic virus* (PSbMV) resistance. Mol Breed 26:425–438

Smýkal P (2006) Development of an efficient retrotransposon-based fingerprinting method for rapid pea variety identification. J Appl Genet 47:221–230

Smýkal P, Coyne C, Redden R, Maxted NP (2013) Chapter 3. In: Singh M, Bisht IS (eds). Genetic and genomic resources for grain legume improvement. London: Elsevier Insights p 41–80

Somers DA, Samac DA, Olhoft PM (2003) Recent advances in legume transformation. Plant Physiol 131:892–899

Srivastava RK, Mishra SK, Singh AK, Mohapatra T (2012) Development of a coupling-phase SCAR marker linked to the powdery mildew resistance gene '*er1*' in pea (*Pisum sativum* L.). Euphytica 186:855–866

Svabova L, Griga M (2008) The effect of cocultivation treatments on transformation efficiency in pea (*Pisum sativum* L.). Plant Cell Tissue Organ Cult 95:293–304

Svabova L, Smýkal P, Griga M, Ondrej V (2005) *Agrobacterium*-mediated transformation of *Pisum sativum in vitro* and *in vivo*. Biol Plantarum 49:361–370

Swiecicki WK, Timmerman-Vaughan G (2005) Molecular marker systems in plant breeding and crop improvement, localization of important traits: the example Pea (*Pisum sativum* L.), Chapter 55. In: Lorz H, Wenzel G (eds) Biotechnology in agriculture and forestry. Springer, Berlin, pp 155–169

Tar'an B, Warkentin T, Somers DJ, Miranda D, Vandenberg A, Blade S, Woods S, Bing D, Xue A, DeKoeyer D, Penner G (2003) Quantitative trait loci for lodging resistance, plant height and partial resistance to mycosphaerella blight in field pea (*Pisum sativum* L.). Theor Appl Genet 107:1482–1491

Tar'an B, Zhang C, Warkentin TD, Tullu A, Vandenberg A (2005) Genetic diversity among varieties and wild species accessions of pea (*Pisum sativum* L.) based on molecular markers, and morphological and physiological characters. Genome 48:257–272

Taylor NL, Heazlewood JL, Day DA, Millar AH (2005) Differential impact of environmental stresses on the pea mitochondrial proteome. Mol Cell Proteomics 4:1122–1133

Timmerman-Vaughan GM, Frew TJ, Butler R, Murray S, Gilpin M, Falloon K, Johnston P, Lakeman MB, Russell A, Khan T (2004) Validation of quantitative trait loci for Ascochyta blight resistance in pea (*Pisum sativum* L.), using populations from two crosses. Theor Appl Genet 109:1620–1631

Timmerman-Vaughan GM, Frew TJ, Russell AC, Khan T, Butler R, Gilpin M, Murray S, Falloon K (2002) QTL mapping of partial resistance to field epidemics of *Ascochyta* blight of peas. Crop Sci 42:2100–2111

Timmerman-Vaughan GM, Frew TJ, Weeden NF (2000) Characterization and linkage mapping of R-gene analogous DNA sequences in pea (*Pisum sativum* L.). Theor Appl Genet 101: 241–247

Timmerman-Vaughan GM, Mc Callum JA, Frew TJ, Weeden NF, Russell AC (1996) Linkage mapping of quantitative trait loci controlling seed weight in pea (*Pisum sativum* L.). Theor Appl Genet 93:431–439

Timmerman-Vaughan G, Larsen R, Murray S, McPhee K, Coyne C (2009) Analysis of the accumulation of Pea enation mosaic virus genomes in seed tissues and lack of evidence for seed transmission in Pea (*Pisum sativum*). Phytopathology 99:1281–1288

Tiwari KR, Penner GA, Warkentin TD (1997) Inheritance of powdery mildew resistance in pea. Can J Plant Sci 77:307–310

Tiwari KR, Penner GA, Warkentin TD (1993) Identification of AFLP markers for powdery mildew resistance gene *er2* in pea. Pisum Genet 31:27–29

Ubayasena L, Bett K, Tar'an B, Vijayan P, Warkentin TD (2010) Genetic control and QTL mapping of cotyledon bleaching resistance in green field pea (*Pisum sativum* L.). Genome 53:346–359

Ubayasena L, Bett K, Tar'an B, Warkentin TD (2011) Genetic control and identification of QTLs associated with visual quality traits of field pea (*Pisum sativum* L.). Genome 54:261–272

Upadhyaya HD, Dwivedi SL, Ambrose M, Ellis N, Berger J, Smýkal P, Debouck D, Duc G, Dumet D, Flavell A, Sharma SK, Mallikarjuna N, Gowda CL (2011) Legume genetic resources: management, diversity assessment, and utilization in crop improvement. Euphytica 180:27–47

Van Der Maesen LJG (1998) Wild plants as genetic resources for crop improvement. In: Mathew P, Sivadasan M (eds) Diversity and taxonomy of tropical flowering plants. Mentor Books, Calicut, pp 93–112

Varhney RK, Chen W, Li Y et al (2012) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotech 30:83–128

Vavilov NI (1922) The law of homologous series in variation. J Genet 12:47–89

Vavilov NI (1949) The phytogeographic basis of plant breeding. In: Chester (Trans.) KS (ed) The origin, variation, immunity, and breeding of cultivated plants. Chronica Botanica, Waltham, MA, pp 13–54

Vershinin AV, Allnutt TR, Knox MR, Ambrose MJ, Ellis NTH (2003) Transposable elements reveal the impact of introgression, rather than transposition, in *Pisum* diversity, evolution, and domestication. Mol Biol Evol 20:2067–2075

Vigeolas H, Chinoy C, Zuther E, Blessington B, Geigenberger P, Domoney C (2008) Combined metabolomic and genetic approaches reveal a link between the polyamine pathway and Albumin 2 in developing pea seeds. Plant Physiol 146:74–82

Vijayalakshmi S, Yadav K, Kushwaha C, Sarode SB, Srivastava CP, Chand R, Singh BD (2005) Identification of RAPD markers linked to the rust (*Uromyces fabae*) resistance gene in pea (*Pisum sativum*). Euphytica 144:265–274

Wang W-Q, Moller IM, Song S-Q (2012) Proteomic analysis of embryonic axis of *Pisum sativum* seeds during germination and identification of proteins associated with loss of desiccation tolerance. J Proteomics 77:68–86. doi:10.1016/j.jprot.2012.07.005

Wang Z, Luo Y, Li X, Wang L, Xu S, Yang J, Weng L, Sato S, Tabata S, Ambrose M, Rameau C, Feng X, Hu X, Luo D (2008) Genetic control of floral zygomorphy in pea (*Pisum sativum* L.). Proc Natl Acad Sci U S A 105:10414–10419

Warkentin TD, Jordan MC, Hobbs LA (1992) Effect of promoter-leader sequences on transient reporter gene expression in particle bombardment pea (*Pisum sativum* L.) tissues. Plant Sci 87:171–177

Weeden NF, Boone WE (1999) Mapping the *Rb* locus on linkage group III using long PCR followed by endonuclease digestion. Pisum Genet 31:36

Weeden NF, Ellis THN, Timmerman-Vaughan GM, Swiecicki WK, Rozov SM, Berdnikov VA (1998) A consensus linkage map for *Pisum sativum*. Pisum Genet 30:1–3

Weeden NF, Marx G (1987) Further genetic analysis and linkage relationships of isozyme loci in the pea: confirmation of the diploid nature of the genome. J Hered 78:153–159

Weeden NF, McGee RJ, Grau CR, Muehlbauer FJ (2000) A gene influencing tolerance to common root rot is located on linkage group IV. Pisum Genet 32:53–55

Weeden NF, Muehlbauer FJ, Ladizinsky G (1992) Extensive conservation of linkage relationships between pea and lentil genetic maps. J Heredity 83:123–129

Weeden NF (2007) Genetic changes accompanying the domestication of *Pisum sativum*: is there a common genetic basis to the 'domestication syndrome' for legumes? Ann Bot 100:1017–1025

Weigelt K, Kuster H, Radchuk R, Muller M, Weichert H, Fait A, Fernie AR, Saalbach I, Weber H (2008) Increasing amino acid supply in pea embryos reveals specific interactions of N and C metabolism, and highlights the importance of mitochondrial metabolism. Plant J 55:909–926

Weigelt K, Küster H, Rutten T, Fait A, Fernie AR, Miersch O, Wasternack C, Emery RJN, Desel C, Hosein F, Müller M, Saalbach I, Weber H (2009) ADP-glucose pyrophosphorylase-deficient pea embryos reveal specific transcriptional and metabolic changes of carbon-nitrogen metabolism and stress responses. Plant Physiol 149:395–411

Weller JL, Hecht V, Liew LC, Sussmilch FC, Wenden B, Knowles CL, Vander Schoor JK (2009) Update on the genetic control of flowering in garden pea. J Exp Bot 60:2493–2499

Weller et al (2012) A conserved molecular basis for photoperiod adaptation in two temperate legumes. Proc Natl Acad Sci USA. doi/10.1073/pnas.1207943110

Wen F, Van Etten HD, Tsaprailis G, Hawe MC (2007) Extracellular proteins in pea root tip and border cell exudates. Plant Physiol 143:773–783

Zhang C, Tar'an B, Warkentin TD, Tullu A, Bett KE, Vandenberg A, Somers DJ (2006) Selection for lodging resistance in field pea (*Pisum sativum* L.) using molecular markers. Crop Sci 46:321–329

Zimmerman J, Saalbach I, Jahn D, Giersberg M, Haehnel S, Wedel J et al (2009) Antibody expressing pea seeds as fodder for prevention of gastrointestinal parasitic infections in chickens. BMC Biotechnol 9:79

Zohary D, Hopf M (2000) Domestication of plants in the old world. Oxford University Press, Oxford

Zong X, Redden R, Liu Q, Wang S, Guan J, Liu J, Xu Y, Liu X, Gu J, Yan L, Ades P, Ford R (2009) Analysis of a diverse global *Pisum* sp. collection and comparison to a Chinese local *P. sativum* collection with microsatellite markers. Theor Appl Genet 118:193–204

Index

S. Gupta et al. (eds.), *Legumes in the Omic Era*, DOI 10.1007/978-1-4614-8370-0,

T