**Compendium of Plant Genomes**
*Series Editor:* Chittaranjan Kole

Rajeev K. Varshney
Manish K. Pandey
Naveen Puppala *Editors*

# The Peanut Genome

Springer

# Compendium of Plant Genomes

**Series editor**

Prof. Chittaranjan Kole, Raja Ramanna Fellow, Department of Atomic Energy, Government of India, Kalyani, India

Whole-genome sequencing is at the cutting edge of life sciences in the new millennium. Since the first genome sequencing of the model plant Arabidopsis thaliana in 2000, whole genomes of about 70 plant species have been sequenced and genome sequences of several other plants are in the pipeline. Research publications on these genome initiatives are scattered on dedicated web sites and in journals with all too brief descriptions. The individual volumes elucidate the background history of the national and international genome initiatives; public and private partners involved; strategies and genomic resources and tools utilized; enumeration on the sequences and their assembly; repetitive sequences; gene annotation and genome duplication. In addition, synteny with other sequences, comparison of gene families and most importantly potential of the genome sequence information for gene pool characterization and genetic improvement of crop plants are described.

**Interested in editing a volume on a crop or model plant?** Please contact Dr. Kole, Series Editor, at ckole2012@gmail.com

More information about this series at http://www.springer.com/series/11805

Rajeev K. Varshney · Manish K. Pandey
Naveen Puppala
Editors

# The Peanut Genome

Springer

*Editors*
Rajeev K. Varshney
Center of Excellence in
Genomics (CEG)
International Crops Research Institute for
the Semi-Arid Tropics (ICRISAT)
Patancheru
India

Manish K. Pandey
Center of Excellence in
Genomics (CEG)
International Crops Research Institute for
the Semi-Arid Tropics (ICRISAT)
Patancheru
India

Naveen Puppala
Agricultural Science Center at Clovis
New Mexico State University
Clovis, NM
USA

ISSN 2199-4781 ISSN 2199-479X (electronic)
Compendium of Plant Genomes
ISBN 978-3-319-63933-8 ISBN 978-3-319-63935-2 (eBook)
DOI 10.1007/978-3-319-63935-2

Library of Congress Control Number: 2017952509



Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

*This book series is dedicated to*
*my wife Phullara, and our children*
*Sourav, and Devleena*

Chittaranjan Kole

# Preface to the Series

Genome sequencing has emerged as the leading discipline in the plant sciences coinciding with the start of the new century. For much of the twentieth century, plant geneticists were only successful in delineating putative chromosomal location, function, and changes in genes indirectly through the use of a number of 'markers' physically linked to them. These included visible or morphological, cytological, protein, and molecular or DNA markers. Among them, the first DNA marker, the RFLPs, introduced a revolutionary change in plant genetics and breeding in the mid-1980s, mainly because of their infinite number and thus potential to cover maximum chromosomal regions, phenotypic neutrality, absence of epistasis, and codominant nature. An array of other hybridization-based markers PCR-based markers, and markers based on both facilitated construction of genetic linkage maps, mapping of genes controlling simply inherited traits and even gene clusters (QTLs) controlling polygenic traits in a large number of model and crop plants. During this period a number of new mapping populations beyond $F_2$ were utilized and a number of computer programs were developed for map construction, mapping of genes, and for mapping of polygenic clusters or QTLs. Molecular markers were also used in studies of evolution and phylogenetic relationship, genetic diversity, DNA-fingerprinting, and map-based cloning. Markers tightly linked to the genes were used in crop improvement employing the so-called marker-assisted selection. These strategies of molecular genetic mapping and molecular breeding made a spectacular impact during the last one and a half decades of the twentieth century. But still they remained "indirect" approaches for elucidation and utilization of plant genomes since much of the chromosomes remained unknown and the complete chemical depiction of them was yet to be unraveled.

Physical mapping of genomes was the obvious consequence that facilitated development of the 'genomic resources' including BAC and YAC libraries to develop physical maps in some plant genomes. Subsequently, integrated genetic-physical maps were also developed in many plants. This led to the concept of structural genomics. Later on, emphasis was laid on EST and transcriptome analysis to decipher the function of the active gene sequences leading to another concept defined as functional genomics. The advent of techniques of bacteriophage gene and DNA sequencing in the 1970s was extended to facilitate sequencing of these genomic resources in the last decade of the twentieth century.

As expected, sequencing of chromosomal regions would have led to too much data to store, characterize, and utilize with the-then available computer software could handle. But development of information technology made the life of biologists easier by leading to a swift and sweet marriage of biology and informatics and a new subject was born—bioinformatics.

Thus, evolution of the concepts, strategies and tools of sequencing and bioinformatics reinforced the subject of genomics—structural and functional. Today, genome sequencing has traveled much beyond biology and involves biophysics, biochemistry and bioinformatics!

Thanks to the efforts of both public and private agencies, genome sequencing strategies are evolving very fast, leading to cheaper, quicker and automated techniques right from clone-by-clone and whole-genome shotgun approaches to a succession of second generation sequencing methods. Development of software of different generations facilitated this genome sequencing. At the same time, newer concepts and strategies were emerging to handle sequencing of the complex genomes, particularly the polyploids.

It became a reality to chemically—and so directly—define plant genomes, popularly called whole-genome sequencing or simply genome sequencing.

The history of plant genome sequencing will always cite the sequencing of the genome of the model plant *Arabidopsis thaliana* in 2000 that was followed by sequencing the genome of the crop and model plant rice in 2002. Since then, the number of sequenced genomes of higher plants has been increasing exponentially, mainly due to the development of cheaper and quicker genomic techniques and, most importantly, development of collaborative platforms such as national and international consortia involving partners from public and/or private agencies.

As I write this preface for the first volume of the new series "Compendium of Plant Genomes", a net search tells me that complete or nearly complete whole-genome sequencing of 45 crop plants, eight crop and model plants, eight model plants, 15 crop progenitors and relatives, and three basal plants are accomplished, the majority of which are in the public domain. This means that we nowadays know many of our model and crop plants chemically, i.e., directly, and we may depict them and utilize them precisely better than ever. Genome sequencing has covered all groups of crop plants. Hence, information on the precise depiction of plant genomes and the scope of their utilization is growing rapidly every day. However, the information is scattered in research articles and review papers in journals and dedicated web pages of the consortia and databases. There is no compilation of plant genomes and the opportunity of using the information in sequence-assisted breeding or further genomic studies. This is the underlying rationale for starting this book series, with each volume dedicated to a particular plant.

Plant genome science has emerged as an important subject in academia, and the present compendium of plant genomes will be highly useful both to students and teaching faculties. Most importantly, research scientists involved in genomics research will have access to systematic deliberations on the plant genomes of their interest. Elucidation of plant genomes is not only of interest for the geneticists and breeders, but also for practitioners of an array of plant science disciplines, such as taxonomy, evolution, cytology,

physiology, pathology, entomology, nematology, crop production, biochemistry, and obviously bioinformatics. It must be mentioned that information regarding each plant genome is ever-growing. The contents of the volumes of this compendium are therefore focusing on the basic aspects of the genomes and their utility. They include information on the academic and/ or economic importance of the plants, description of their genomes from a molecular genetic and cytogenetic point of view, and the genomic resources developed. Detailed deliberations focus on the background history of the national and international genome initiatives, public and private partners involved, strategies and genomic resources and tools utilized, enumeration on the sequences and their assembly, repetitive sequences, gene annotation, and genome duplication. In addition, synteny with other sequences, comparison of gene families, and, most importantly, potential of the genome sequence information for gene pool characterization through genotyping by sequencing (GBS) and genetic improvement of crop plants have been described. As expected, there is a lot of variation of these topics in the volumes based on the information available on the crop, model, or reference plants.

I must confess that as the series editor it has been a daunting task for me to work on such a huge and broad knowledge base that spans so many diverse plant species. However, pioneering scientists with life-time experience and expertise on the particular crops did excellent jobs editing the respective volumes. I myself have been a small science worker on plant genomes since the mid-1980s and that provided me the opportunity to personally know several stalwarts of plant genomics from all over the globe. Most, if not all, of the volume editors are my longtime friends and colleagues. It has been highly comfortable and enriching for me to work with them on this book series. To be honest, while working on this series I have been and will remain a student first, a science worker second, and a series editor last. And I must express my gratitude to the volume editors and the chapter authors for providing me the opportunity to work with them on this compendium.

I also wish to mention here my thanks and gratitude to the Springer staff, Dr. Christina Eckey and Dr. Jutta Lindenborn in particular, for all their constant and cordial support right from the inception of the idea.

I always had to set aside additional hours to edit books besides my professional and personal commitments—hours I could and should have given to my wife, Phullara, and our kids, Sourav, and Devleena. I must mention that they not only allowed me the freedom to take away those hours from them but also offered their support in the editing job itself. I am really not sure whether my dedication of this compendium to them will suffice to do justice to their sacrifices for the interest of science and the science community.

Kalyani, India Chittaranjan Kole

# Preface

Genome contains a set of genetic instructions coded in the form of just four letters (A, C, G, T) which defines the basic behavior of every plant species. Since the discovery of DNA, researchers have continuously been trying to understand the instructions encoded in the genome and finding out ways to manipulate these instructions for achieving desirable phenotype in a crop species. The pace of such understanding for desirable traits in peanut has been extensively slow because of the genetic complexity and large-sized genome. The availability of reference genome sequences for both the diploid progenitor species has provided acceleration to this process of understanding and deploying modern approaches for candidate gene discovery and marker development for key traits in peanut. Genomics research has revolutionized the pace of genetics and breeding research due to low-cost sequencing and high throughput genotyping technologies. These resources not only helped in developing better understanding the basic biology of the crop plants but also used together with other genetic resources for developing genomics tools to deploy them in breeding for developing improved varieties.

The peanut, also known as groundnut (*Arachis hypogaea*), is an important legume crop mainly utilized for cooking oil and confectionary and table purpose. This crop is widely cultivated in >100 countries with a total production of million tons during 2014. The cultivated peanut came into existence from hybridization between two diploid species (*A. duranensis* and *A. ipaensis*) possessing different genomes. The current understanding is that the hybridization event gave rise initially to a wild form of tetraploid peanut species, *A. monticola*, which after the long domestication process gave rise to the cultivated tetraploid species, dramatically different from its wild relatives. This crop has unique feature of geotropism and skotomorphogenesis, i.e., flowering happens above ground and seed development happens below the ground. The year 2016 has been a very significant year for peanut research community as reference genome sequence for both the diploid progenitors as well as a high throughput genotyping array with 56K single nucleotide polymorphisms (SNPs) have become available for genomics studies, candidate gene discovery, high resolution trait-mapping, and marker development and breeding.

This book is very timely in peanut as part of the genome compendium series for different crops. It contains 11 different chapters providing detailed overview on different aspects of botanical classification, genetics, genomics,

and breeding. This book not only provides information on recent advances on genome sequencing, genome architecture, genetic mapping for few traits and marker identification, but also presents case studies of developing molecular breeding products for foliar diseases, nematode resistance, and oil quality.

A total of 30 authors from Argentina, Brazil, China, India, and USA have contributed 11 chapters for this volume (see "Contributors"). The editors of this volume are grateful to all the authors for their contribution in writing chapters of high quality of their area of expertise and reviewers (see "Reviewers") for their constructive suggestions and corrections helping in improving the quality of the chapters further. The editors are also thankful to Dr. David Bergvinson, Director General, ICRISAT and Dr. Peter Carberry, Deputy Director General—Research, ICRISAT for their support. The editors thank Prof. C. Kole, Series Editor for his invitation and help in editing this volume. The cooperation received from Abirami Purushothaman, Jegadeeswari Diravidamani and Nareshkumar Mani from Springer has been a great help in completion of this book and is gratefully acknowledged. The cooperation and encouragement from publisher have been of great help in completion of this book and are gratefully acknowledged.

In addition to above, we also appreciate and recognize cooperation and moral support from our family members for sparing us precious time for editorial work that we should have spent with our respective families. RKV acknowledges the help and support of wife (Monika), son (Prakhar), and daughter (Preksha) who allowed their time to be taken away to fulfill RKV's editorial responsibilities in addition to research and other administrative duties at ICRISAT. Similarly, MKP is grateful to his wife (Seema) for her help and moral support while doing editorial responsibilities in addition to research duties at ICRISAT with special thanks to his brave daughter (late Tanisha) who has been alive for only a short period of time (3 months) after birth. NP also acknowledges his wife (Vani) and son (Kunal) for their cooperation and understanding in not fulfilling the family responsibilities during the evenings and weekends due to editorial commitment.

We hope that our efforts in compiling the information on different aspects of peanut will help the peanut genomics and breeding researchers in developing better understanding and research strategies. This book will also benefit students, academicians, and policy makers in updating their knowledge on recent advances in peanut research.

Patancheru, India — Rajeev K. Varshney
Patancheru, India — Manish K. Pandey
Clovis, USA — Naveen Puppala

# Contents

# Contributors

**A.C.G. Araujo** Embrapa Genetic Resources and Biotechnology (Cenargen) - Parque Estação Biológica (PqEB), Brasília, DF, Brazil

**A.C.M. Brasileiro** Embrapa Genetic Resources and Biotechnology (Cenargen) - Parque Estação Biológica (PqEB), Brasília, DF, Brazil

**Ethalinda K.S. Cannon** Department of Computer Science, Iowa State University, Ames, IA, USA

**Steven B. Cannon** Corn Insects and Crop Genetics Research Unit, US Department of Agriculture–Agricultural Research Service, Ames, IA, USA

**Laura Chalup** Instituto de Botánica del Nordeste (CONICET-UNNE), Corrientes, Argentina

**Patricia M. Guimarães** Embrapa Genetic Resources and Biotechnology (Cenargen) - Parque Estação Biológica (PqEB), Brasília, DF, Brazil

**Baozhu Guo** USDA-ARS, Crop Protection and Management Research Unit, Tifton, USA

**Xiaohong Guo** USDA-ARS, Tifton, GA, USA; University of Georgia, Tifton, GA, USA; Heilongjiang Bayi Agricultural University, Daqing, China

**Scott A. Jackson** Center for Applied Genetic Technologies, University of Georgia (UGA), Athens, USA

**P. Janila** International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Telangana, India

**Xiangyun Ji** USDA-ARS, Tifton, GA, USA; University of Georgia, Tifton, GA, USA; Ecological Environment Protection Research Institute, Shanghai Academy of Agricultural Sciences, Shanghai, China

**Graciela I. Lavia** Instituto de Botánica del Nordeste (CONICET-UNNE), Corrientes, Argentina; Facultad de Ciencias Exactas y Naturales y Agrimensura, Universidad Nacional del Nordeste, Corrientes, Argentina

**Xuanqiang Liang** Crop Research Institute, Guangdong Academy of Agricultural Sciences (GAAS), Guangzhou, China

**Boshou Liao** Oil Crops Research Institute (OCRI) of Chinese Academy of Agricultural Sciences (CAAS), Wuhan, China

**A. Mehta** Embrapa Genetic Resources and Biotechnology (Cenargen) - Parque Estação Biológica (PqEB), Brasília, DF, Brazil

**Spurthi N. Nayak** International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India

**Victor Nwosu** Plant Science Program Research and Development, MARS Chocolate North America, Hackettstown, NJ, USA

**Alejandra M. Ortiz** Instituto de Botánica del Nordeste (CONICET-UNNE), Corrientes, Argentina; Facultad de Ciencias Exactas y Naturales y Agrimensura, Universidad Nacional del Nordeste, Corrientes, Argentina

**Peggy Ozias-Akins** Institute of Plant Breeding, Genetics and Genomics, University of Georgia, Tifton, GA, USA

**Manish K. Pandey** International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India

**Naveen Puppala** New Mexico State University, Agricultural Science Center, Clovis, USA

**Germán Robledo** Instituto de Botánica del Nordeste (CONICET-UNNE), Corrientes, Argentina; Facultad de Ciencias Exactas y Naturales y Agrimensura, Universidad Nacional del Nordeste, Corrientes, Argentina

**Sergio S. Samoluk** Instituto de Botánica del Nordeste (CONICET-UNNE), Corrientes, Argentina

**Guillermo Seijo** Instituto de Botánica del Nordeste (CONICET-UNNE), Corrientes, Argentina; Facultad de Ciencias Exactas y Naturales y Agrimensura, Universidad Nacional del Nordeste, Corrientes, Argentina

**María C. Silvestri** Instituto de Botánica del Nordeste (CONICET-UNNE), Corrientes, Argentina

**Shyam P. Tallury** Plant Genetic Resources Conservation Unit, USDA-ARS, Griffin, GA, USA

**Murali T. Variath** International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Telangana, India

**Rajeev K. Varshney** International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India; University of Western Australia (UWA), Crawley, Australia

**Manish K. Vishwakarma** International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India

**Liyun Wan** Oil Crops Research Institute (OCRI) of the Chinese Academy of Agricultural Sciences, Wuhan, China

**Hui Wang** USDA-ARS, Tifton, GA, USA; University of Georgia, Tifton, GA, USA

# Reviewers

**Bera, S.K.**, ICAR-Directorate of Groundnut Research (DGR), Junagadh, India

**Bhat, Ramesh**, University of Agricultural Sciences (UAS), Dharwad, India

**Chen, Xiaoping**, Crop Research Institute (CRI), Guangdong Academy of Agricultural Sciences (GAAS), Guangzhou, China

**Dutta, Manoranjan**, ICAR-National Bureau of Plant Genetic Resources (NBPGR), New Delhi, India

**Guo, Baozhu**, Crop Protection and Management Research Unit (CPMRU), US Department of Agriculture–Agricultural Research Service (USDA-ARS), Tifton, Georgia, USA

**Janila, P.**, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India

**Radhakrishnan, T.**, ICAR-Directorate of Groundnut Research (DGR), Junagadh, India

**Rathnakumar, A.L.**, ICAR-Directorate of Groundnut Research (DGR), Junagadh, India

# The Peanut Genome: An Introduction

1

Rajeev K. Varshney, Manish K. Pandey and Naveen Puppala

**Abstract**

Peanut is an important oilseed and legume crop with global relevance to food and nutritional security in addition to source of income to millions of stallholder farmers of the Sub-Saharan Africa and Asia. Low genetic diversity in cultivated genepool and ploidy differences between different genepools have been the two important genetic bottlenecks hampering use of molecular breeding approaches for peanut improvement. Nevertheless, recent advances in genomics research have elevated the status of peanut from "resource-poor" to "genomic resource-rich" crop and therefore, it is an obligation to the peanut research community across the world to adopt a holistic approach including use of genomics information and tools in crop improvement programs. In this context, this book provides up-to-date information on the progress made in last 5 years in peanut genomics with a particular focus on the latest genomic findings, tools, and strategies employed in genome sequencing, transcriptomics, trait mapping, and molecular breeding approaches. This chapter by providing an overview of the contents of the book presents a big picture on the current status of peanut genome and allied information and its potential applications for peanut improvement.

## 1.1 Introduction

Peanut (*Arachis hypogaea* L.), also known as groundnut, is a global crop being cultivated in >100 countries while consumed by almost all the human societies across the world. Currently, peanut is grown in 26.5 million hectares (Mha) globally which yields 43.9 million tons (Mt) with the productivity rate of 1.65 tons/hectare (t/ha) (http://www.fao.org/

R.K. Varshney (✉) · M.K. Pandey
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India
e-mail: r.k.varshney@cgiar.org

R.K. Varshney
University of Western Australia (UWA), Crawley, Australia

N. Puppala
New Mexico State University, Agricultural Science Center, Clovis, USA


R.K. Varshney et al. (eds.), *The Peanut Genome*, Compendium of Plant Genomes, DOI 10.1007/978-3-319-63935-2_1

faostat/en/#data/QC). There is a huge difference in productivity between Americas (3.3 t/ha), Asia (2.4 t/ha) and Africa (0.96 t/ha) regions (Fig. 1.1). It is important to note that the productivity rate in Asia is better than Africa because of higher productivity achieved in China, while India's productivity is close to the global productivity rate. In other words, Africa produces mere 31.5% of global produce from 54.3% peanut growing area, while Asia produces just 58.3% of global production from 40.7% area in contrast to Americas which produces 10.0% of global produce just from 4.9% of global peanut-producing area (Table 1.1).

This crop is consumed in multiple forms (roasted seeds/oil/confectionary) throughout the world and has different significance to different regions of the world. For example, this crop only serves the table purpose/confectionary in Americas; as vegetable oil and confectionary/table purpose in Asia; while as nutritional supplement and confectionary/table purpose in Africa (Pandey et al. 2012; Janila et al. 2016). Further, this crop matters a lot to the resource-poor farmers in Africa by contributing significantly toward achieving food and nutritional security in addition to the financial security through income generation. Further, the limiting natural resources (land, water, genetic diversity, etc.,), unsolved existing biotic and abiotic stress and expected climate change problems in future will make the conditions for plants even difficult than ever. The highest reduction in yield and quality of the produce may occur in the fields of small-holder and resource-poor farmers in the developing countries. Therefore, it is an obligation to the peanut research community across the world to adopt a holistic approach including the use of genomics information and tools in the crop improvement programs to develop climate-smart peanut varieties that can help in improving livelihood of poor farmers in developing countries.

The last decade has witnessed path-breaking discoveries using next-generation sequencing and genomics technologies in several crops. Genomics is becoming an integral part of breeding programs to facilitate accelerated development of improved varieties. Peanut has also witnessed significant development (especially in last 5 years) including sequencing genome of both the diploid progenitors, huge transcriptome resources, large-scale genomic variations to use as genetic markers, genetic populations (bi- and multi-parent populations and germplasm sets), marker-trait associations, and molecular breeding products (Pandey et al. 2016). The immediate availability of genome sequence for tetraploid cultivated peanut will be the most useful genomic resource for better understanding of traits and use in breeding program. It will be interesting to see the greater and effective role of genomics information in transferring the superior alleles from wild species/un-adapted germplasm to elite varieties in order to rule out the inherent problem of linkage drag (Varshney et al. 2013). Therefore, it is the high time to adopt and deploy integrated breeding approach wherein the information on genomics, proteomics, bioinformatics, and phenomics will be used for breeding improved peanut varieties which can withstand in the farmer's field. In view of above, Editors planned this book to provide one-stop shop for providing all information related to peanut genome and its application for crop improvement. This chapter provides a summary of different chapters included in the book under the following sections.

**Table 1.1** Current global peanut cultivated area, production, and productivity scenario

| | Area (Mha) | Production (Mt) | Productivity (Kg/ha) |
|---|---|---|---|
| World | 26.5 | 43.9 | 1655 |
| Africa | 14.4 (54.3%) | 13.9 (31.7%) | 965 |
| Asia | 10.8 (40.7%) | 25.6 (58.3%) | 2371 |
| Americas | 1.3 (4.9%) | 4.4 (10.0%) | 3333 |

% indicates against global peanut cultivation area and production by Africa, Asia, and Americas

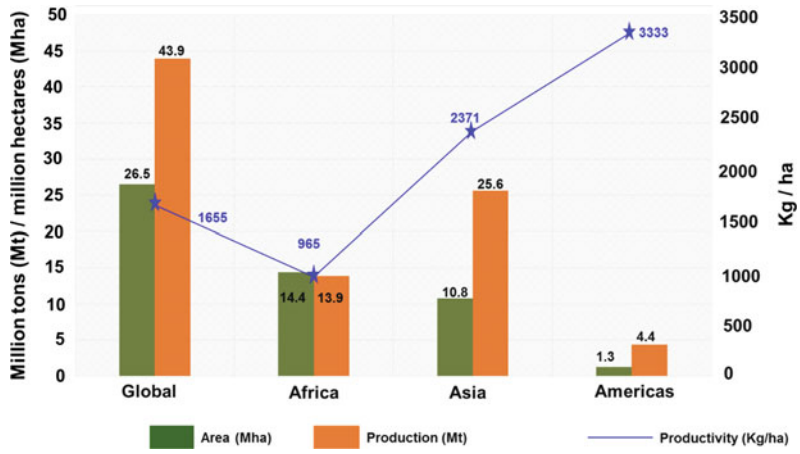


**Fig. 1.1** Peanut area, production and productivity in different continents of the world during 2014

## 1.2 Importance, Botanical Description, and Cytogenetics

Chapter 2 entitled "Economic and Academic Importance of Peanut" provides detailed information on current global production, its yield constraints and use of different plant parts (plant as fodder, kernel, skin, shell, cake, and roots). This chapter also discusses on different aspects of demand for peanut and peanut-based products in the international market, especially for confectionary types. The trend of peanut consumption is going toward confectionary and table purpose as there is a sharp demand for diversified peanut products as confectionary and table purpose in addition to competition from other sources of vegetable oils available in the market. Further, aflatoxin contamination poses a very serious concern in international market in the selling produce to the countries across the globe having very high stringent quality criteria. The chapter also provides enough information which proves peanut a rich source of nutrition and provides several health benefits. Further information were also provided on botanical classification and botanical types of cultivated peanuts along with current status on developing improved varieties across different breeding programs of the world.

Chapter 3, "Origin and Botanical Descriptions of Peanut", authored by Shyam Tallury from USDA-ARS, Griffin (USA) provides information on the genetic structure of the genus including its origin, variability, and geographical distribution of various species. This chapter emphasizes the importance of preserving the genetic diversity and characterization of the germplasm resources for efficient utilization in peanut improvement programs. The origin and botanical classification of the cultivated species have been revisited in the chapter to summarize our current knowledge of the taxon. This chapter reports the existence of few taxonomic and phylogenetic ambiguities, and provides hope for better clarification and understanding of the genetic structure of genus *Arachis* and cultivated species, *A. hypogaea* using the recent advances in peanut genome sequencing and the availability of improved genomics tools.

Chapter 4 entitled "Cytological Features of Peanut Genome" authored by Guillermo Seijo and colleagues provides detailed information on the chromosomal features of peanut by using classical and molecular cytogenetic technologies. This chapter discusses in detail about different karyotyping studies to unravel the taxonomy of the genus and to establish relationships among species. Critical information on chromosome morphology, heterochromatin, and genome sizes of *Arachis* species has also been presented to understand origin of cultivated peanut (*A. hypogaea*) from *A. monticola,* a direct tetraploid

antecessor of peanut. The cytogenetic studies not only clearly proved that *A. duranensis* (A genome) and *A. ipaënsis* (B genome) are the diploid progenitors of the AABB tetraploids, but also established the relations among other genome groups (A, B, D, F, G, and K).

## 1.3 Germplasm and Genomics Resources for Trait Mapping

Boshou Liao from Oil Crops Research Institute of Chinese Academy of Agricultural Sciences (CAAS) provides information on the conservation and characterization of peanut germplasm in Chap. 5, "Germplasm Characterization and Trait Discovery in Peanut". The chapter discussed development of core and minicore collections for better utilization of diverse germplasm in routine breeding program. The chapter also mentioned research priorities for important traits such as yield, resistance to biotic stress (late leaf spot, early leaf spot, rust, tomato spotted wilt virus, groundnut rosette virus, bacterial wilt, nematodes, and aflatoxin contamination), tolerance to drought stress, and oil quality. Further deployment of modern genomics approaches has been suggested for accelerated trait/gene discovery and development of appropriate genomics tools for key traits to deploy them in routine breeding program to utilize germplasm collection across the globe.

Peggy Ozias Akins from University of Georgia, Athens (USA) and her collaborators from USDA-ARS, Ames (USA) provide detailed information on different types of genomic resources in Chap. 6 ("Genomics Resources for Peanut Improvement"). The chapter deals with genome sequence of diploid ancestors and soon to be available tetraploid genome along with huge transcriptome sequences, resequencing data for cultivated and wild genotypes, and marker-trait associations for use in breeding. The availability of a large number of structural variations facilitated development of high throughput genotyping array with 58 K SNPs to conduct multiple genetic and breeding studies (Pandey et al. 2017; Clevenger et al. 2017). This chapter also presents on use of publicly available reference genome sequence of peanut's diploid progenitors (peanutbase.org) for enabling expansion of genotyping tools for genetic mapping to facilitate gene-based selection in breeding.

Chapter 7 entitled "Classical and Molecular Approaches for Mapping of Genes and Quantitative Trait Loci in Peanut" authored by scientists from ICRISAT (India), USDA-ARS Tifton (USA), and OCRI-CAAS (China) provides updates on cytological studies, molecular markers, genetic linkage maps, and trait-linked QTL identification using linkage and association mapping/linkage disequilibrium mapping approaches. This chapter also emphasizes on development and deployment of next-generation genetic populations such as MAGIC (multi-parent advanced generation intercross) and NAM (nested association mapping). Further information is provided on different strategies for NGS-based SNPs identification linked to gene/QTLs for concerned traits using modern high-resolution trait mapping and gene discovery approaches. Also, examples have been provided for successful development and deployment of diagnostic markers for improving selected traits in peanut.

## 1.4 Genome Sequence and Beyond

Although rice was the first crop genome to be sequenced in 2002 (Goff et al. 2002), it took 14 years to have the genome sequence of peanut. This can be attributed to the complexity of genome as well as small community of peanut researchers. Therefore, several scientists from China, USA, and ICRISAT involved in sequencing projects present "History and Impact of the International Peanut Genome Initiative: the Exciting Journey toward Peanut Whole-genome Sequencing" in Chap. 8. Authors take readers to a nice journey from the launch of U.S. Peanut Genome Initiative (PGI) in 2004 to the currently functioning International Peanut Genome Initiative (IPGI) that floated Peanut Genome Consortium (PGC). The IPGI initiated the peanut genome sequencing project in 2010 leading to availability of the genome sequences of two

diploid peanut progenitors in 2014 and were published in 2016. This chapter reviewed the background and history of IPGI, its accomplishments in sequencing of peanut as well as advancing of peanut genomics research including sequencing of tetraploid peanut genome.

In Chapter 9 entitled "Genome Sequencing of Ancestor Diploid Genomes for Enhancing Genome Understanding and Peanut Improvement", scientists from ICRISAT, USA, and China present in detail about the approaches and technical details of two genome sequencing initiatives for developing genome assemblies for diploid progenitors (A genome—*Arachis duranensis* and B genome—*A. ipaensis*). The IPGI reported the sequencing of both A and B genomes, while Diploid Progenitor Peanut Arachis Genome Sequencing Consortium (DPPAGSC) reported the sequencing of only A genome. This chapter provides information generated by these two efforts on the genome architecture, organization, size, genes, geocarpy, oil biosynthesis, and allergens besides providing information about evolution and polyploidization.

Besides genome sequencing, it is essential to identify the candidate genes for different traits. Chapter 10 entitled "Functional Genomics in Peanut Wild Relatives" by Patricia Guimarães and her colleagues from Embrapa Genetic Resources and Biotechnology (Cenargen), Brazil presents functional genomics approaches by using peanut wild relatives as these species are rich source of desirable traits including disease resistance that is not available in cultivated genepool. The chapter summarizes information on different studies related to functional genomics including ESTs, unigenes, full-length cDNA clones, and proteins. In recent years, unprecedented amounts of genomic information for wild and cultivated *Arachis* have become available leading to the discovery of genes and regulatory sequences, and enlarging the collections of molecular markers. To maximize these valuable assets, further studies on candidate gene identification and validation are required to find out the functional variation which can be deployed to develop improved varieties using modern science technologies such as genome editing, finding a way out to avoid the genetic barriers of ploidy and genome differences.

Finally, Editors of the book present the opportunities and approaches for peanut improvement in Chap. 11 entitled "Future Prospects for Peanut Improvement". The chapter emphasizes the importance and opportunities arisen due to availability of genome sequences, resequencing, millions of structural variations as genetic markers, next-generation genotyping platforms, and complex multi-parental genetic populations. These genomic and genetic resources will facilitate faster discovery of candidate genes and development of diagnostic markers to breed improved varieties by pyramiding desirable alleles for multiple traits. In addition to marker-assisted backcrossing, the other modern breeding approaches such as genomic selection need more attention to reap benefits for achieving higher genetic gain by improving complex traits such as yield under rainfed conditions.

## 1.5 Conclusion

This book provided an up-to-date information about peanut genome and its improvement by using modern genomics and breeding approaches by a panel of expert and eminent peanut scientists across the world. This book does not just provide the current landscape of peanut genomics at international level in terms of tools and strategies employed in genome sequencing, transcriptomics, functional analysis, trait mapping and molecular breeding but also potential approaches for accelerating genetics and genomics research and enhancing genetic gains in crop improvement. Editors are very much hopeful that this book will be helpful to the peanut research community to develop a road map for deploying genomics technologies for peanut improvement.

## References

Clevenger J, Chu Y, Chavarro C, Agarwal G, Bertioli DJ, Leal-Bertioli SCM, Pandey MK, Vaughn J, Abernathy B, Barkley N, Hovav R, Burow M, Nayak SN, Chitikineni A, Isleib T, Holbrook C, Jackson SA, Varshney RK, Ozias-Akins P (2017) Genome-wide SNP genotyping resolves signatures of selection and tetrasomic recombination in peanut. Mol Plant 10:309–322

Goff SA, Ricke D, Lan TH et al (2002) A draft sequence of the rice genome (*Oryza sativa L. ssp. japonica*). Science 296(5565):92–100

Janila P, Variath MT, Pandey MK, Desmae H, Motagi B, Okori P, Manohar SS, Rathnakumar AL, Radhakrishnan T, Boshou Liao, Varshney Rajeev K (2016) Genomic tools in groundnut breeding program: status and perspectives. Frontiers in Plant Science 7:289. doi:10.3389/fpls.2016.00289

Pandey MK, Monyo E, Ozias-Akins P, Liang X, Guimarães P, Nigam SN, Upadhyaya HD, Janila P, Zhang X, Guo B, Cook DR, Bertioli DJ, Michelmore R, Varshney RK (2012) Advances in *Arachis* genomics for peanut improvement. Biotechnol Adv 30 (3):639–651

Pandey MK, Roorkiwal M, Singh V, Lingam A, Kudapa H, Thudi M, Chitikineni A, Rathore A, Varshney RK (2016) Emerging genomic tools for legume breeding: current status and future perspectives. Frontiers in Plant Sci 7:455. doi:10.3389/fpls.2016.00455

Pandey MK, Agarwal G, Kale SM, Clevenger J, Nayak SN, Sriswathi M, Chitikineni A, Chavarro C, Chen X, Upadhyaya HD, Vishwakarma MK, Leal-Bertioli S, Liang X, Bertioli DJ, Guo B, Jackson SA, Ozias-Akins P, Varshney RK (2017) Development and evaluation of a high density genotyping 'Axiom_Arachis' array with 58K SNPs for accelerating genetics and breeding in groundnut. Sci Rep 7:40577

Varshney RK, Mohan SM, Gaur PM, Gangarao NVPR, Pandey MK, Bohra A, Sawargaonkar S, Kimurto PK, Janila P, Saxena KB, Fikre A, Sharma M, Pratap A, Tripathi, S Datta, SK Chaturvedi, G Anuradha, A Babbar, RG Chaudhary, MB Mhase, CH Bharadwaj S, Mannur DM, Harer PN, Guo B, Liang X, Nadarajan N, Gowda CLL (2013) Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. Biotechnol Adv 31:1120–1134

# Economic and Academic Importance of Peanut

2

Murali T. Variath and P. Janila

**Abstract**

Peanut is an important oil, food and feed crop of the world. The kernels are rich in fats and protein, and 100 g of kernels provide 567 kcal of energy and 8.5 g of dietary fiber. Peanuts are source of minerals, vitamins and antioxidants and health improving bioactive compounds such as resveratrol, tocopherol, arginine etc. and hence are touted as functional food. Consumption of peanuts can reduce risk of inflammation, diabetes, cancer, alzheimer's and gallstone disease. Peanut is cultivated in over 100 countries, with over 95% of cultivated area in Asia and Africa. Aflatoxin and allergens are major health deterrents in peanut and more research efforts are needed to develop aflatoxin and allergen free peanuts. There is a great demand for peanut and peanut-based products in the international market, especially for confectionary types. Breeding new cultivars that meet the needs of the producers, consumers and industry is an important research area with implications along the value chain. Conventional breeding approaches and phenotyping tools were widely used to breed several varieties and in the last decade, genomic tools are integrated for making selections. The advent of next-generation sequencing (NGS) tools and the availability of the draft genome sequence of the diploid progenitors of peanut *A. duranensis* and *A. ipaensis* is expected to play a key role in sequencing the genome of cultivated peanut. Transgenic peanuts with resistance to herbicide, fungus, virus, and insects; tolerance to drought and salinity and improved grain quality are under testing at different containment levels. The availability of sophisticated tools for both genotyping and phenotyping will lead to an increase in our understanding of key genes involved and their metabolic regulatory pathways.

M.T. Variath · P. Janila (✉)
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Telangana, India
e-mail: p.janila@cgiar.org



R.K. Varshney et al. (eds.), *The Peanut Genome*,
Compendium of Plant Genomes, DOI 10.1007/978-3-319-63935-2_2

## 2.1 Introduction

### 2.1.1 Name, Nomenclature, and Uses of Different Plant Parts

Peanut or groundnut is an economically important oilseed, feed, and food crop and widely cultivated in tropical and subtropical regions of the world. It is known by many local names such as earthnut, goober pea, pindas, jack nut and monkey nut. The botanical name for peanut, *Arachis hypogaea* Linn. is derived from the Greek words 'arachos' meaning a weed and *hypogaea* meaning below ground/underground chamber, referring to a weed/plant that produces fruit/pods in the soil. It belongs to the Leguminosae family, tribe *Aeschynomeneae*, and subtribe *Stylosanthenae*. It is an annual herbaceous plant, growing 30–50 cm tall and bearing tetrafoliate leaves in a 2/5 phyllotaxy (two opposite pairs; no terminal leaflet). Although peanut pods/kernels are the most important product of peanut plant, all parts of the plant are useful and can be utilized in a variety of ways.

**Plant:** The dried haulms after harvest constitute an energy rich fodder for cattle or in manure. A few species of peanut such as *A. glabrata* and *A. pintoi* are cultivated especially for forage purpose in South America, North America, and Australia. Another species *A. repens* is being utilized as a ground cover in South America.

**Seed (kernel):** The peanut seeds are consumed directly as raw, roasted, and boiled or processed into confections and peanut flour for flavor enhancement or crushed for oil for edible and industrial uses. It is also widely used in the preparation of ready to use therapeutic and supplementary foods (RUTF and RUSF) to combat malnutrition in developing countries. The peanut kernels are a source of high-quality edible oil (44–56%), easily digestible protein (22–30%), carbohydrates (10–25%), vitamins (E, K, and B complex), minerals (Ca, P, Mg, Zn and Fe) and fiber. The inferior quality oil has a variety of industrial uses. Paint, varnish, lubricants, leather dressings, furniture polish, insecticides, nitroglycerine, soap, and cosmetics are all prepared from the oil. The protein portion of the oil is utilized in the manufacture of some textile fibers (Bell 2008).

**Peanut skin**: Peanut skins are obtained from processed nuts, broken nuts and sometimes from nuts not found fit for human consumption. They are a good source of several bioactive compounds mainly phenolics and have long been used in China as a traditional Chinese medicine for the treatment of chronic hemorrhage and bronchitis. They are also added as supplements to processed food products such as peanut paste and peanut butter to improve flavor, stability and antioxidant capacity (Hathorn and Sanders 2012).

**Peanut cake:** This byproduct of oil extraction is used in animal feed industry, in making weaning foods for children, invalid foods for aged people and as fertilizer.

**Shell:** The shells are used as fuel, animal feed, cattle litter, filler in feed and fertilizer industry and in making particle boards, and alcohol and acetone after fermentation. They are also used to make cellulose (used in rayon and paper) and mucilage (glue) (Bell 2008).

**Roots:** Being a legume crop the roots add nitrogen (100–152 kg/ha N) and organic matter to the soil (Nigam 2015).

### 2.1.2 Area, Production, and Growing Regions

Peanut is cultivated in 25.70 million ha world over with a total production of 42.32 million tons of in-shell peanuts during 2014 (FAOSTAT 2015). China (22%), India (19%), Nigeria (11%), and the USA (2%) are the major peanut growing countries. China (42%) and India (18%) account for about 60% of the total production globally followed by Nigeria (7.7%), the USA (4.3%) and Indonesia (1.8%) (Rathnakumar et al. 2013). Africa with 12.40 m ha area and 11.54 m tons of production, and Asia with 11.87 m ha and 29.95 m tons, together account for 95% global peanut area and 91% of global peanut production. Since 1980, the global peanut production increased by 2.67% annually until 2014 and this increase was contributed by an increase in both

cultivated area (0.93%) and yield (1.74%) (FAOSTAT 2015). In Asia, the annual growth rate in terms of area cultivated and production increased by 0.05% and 2.60% during the same period while that of Africa was much higher with 2.46% and 3.62% increase in area and production. In terms of peanut oil production, China with 2.74 m tons in 2015/16 was the top producer followed by India (1.1 m tons) and Myanmar (0.27 m tons). India's share in global production of peanut oil is expected to be around 20% in 2015–2016.[1]

Peanut is cultivated in more than 100 countries across the world under different agroecological environments. However, the major cultivation is confined to the developing countries of Asia and Africa which accounts for a bulk of the total peanut area as well as production. The share of Asia in global peanut area and production in 1990 was 65.70 and 70.52% respectively while in 2014 it was 43.62 and 59.79% respectively, indicating a declining trend (Fig. 2.1). In comparison, the African share increased from 27.68 and 19.31% in 1990 to 51.01 and 29.79% respectively in 2014 (Fig. 2.2). China in East Asia, India in South Asia and Myanmar, Indonesia, Vietnam, and Thailand in Southeast Asia are the major Asian peanut producing regions.

### 2.1.3 Yield and Production Constraints

Peanut yields increased worldwide by 1.74% per annum between 1980 and 2014. In 2014 the world average productivity of in-shell peanuts was 1648 kg/ha which was lower as compared to 1823 kg/ha in 2013. Cyprus, Israel, Barbados, Nicaragua, and USA were the top 5 highest yielders during 2013–2014 (FAOSTAT 2015). In the Asian context, China was the top yielder in 2014 with an annual yield of 3490 kg/ha, as compared to 1261 kg/ha of India. During the period of 1980–2014, China showed 4.71% positive annual growth rate while it was 1.40% for India (FAOSTAT 2015). Vietnam (145%), Indonesia (33%), Thailand (32%) and Myanmar (81%) in Southeast Asia experienced spectacular increase in yield during 1981–1983 to 2012–2014 and most of this increase was contributed by the introduction of high-yielding, stress-resistant varieties and improved production practices such as balanced fertilization, efficient weed and chemical pest control, use of polythene mulches and improved technology dispersal systems. The yield increase in China and India during the same period was 116 and 53%, respectively. As compared to regional variation, considerable yield difference is also observed between developed and developing countries which are mainly due to the different production systems being practiced. Peanut is grown in two distinct production systems—low-input system practiced by the farmers of Africa and South Asia and the high-input production systems. In the low-input system, farmers cultivate low-yielding, late-maturing varieties on marginal lands with no irrigation and limited inputs with average yields ranging from 800 to 1000 kg/ha. In high-input production systems practiced in the USA, Australia, Argentina, Brazil, China and South Africa, peanut yields can vary from 2 to 4 t/ha. Here cultivation of peanut is commercialized using improved varieties, modern crop management practices, irrigation and high resource inputs along with fully mechanized farming.

Productivity levels of peanut in most of the developing countries have remained low due to several production constraints which include biotic and abiotic stresses, lack of efficient seed dispersal systems, technological knowhow, market accessibility, low-input use and factors related to socio-economic infrastructure. Insect pests, diseases, drought, and low soil fertility are the major biotic and abiotic stresses. Most of the peanut cultivation in the developing countries is carried out by smallholder farmers who lack adequate resources or access to novel technologies to improve productivity. Additionally, limited market access and low producer prices reduce the incentive for producers to invest in productivity-enhancing technologies such as

[1] http://www.agricoop.nic.in/Admin_Agricoop/Uploaded_File/edib_2201.pdf.

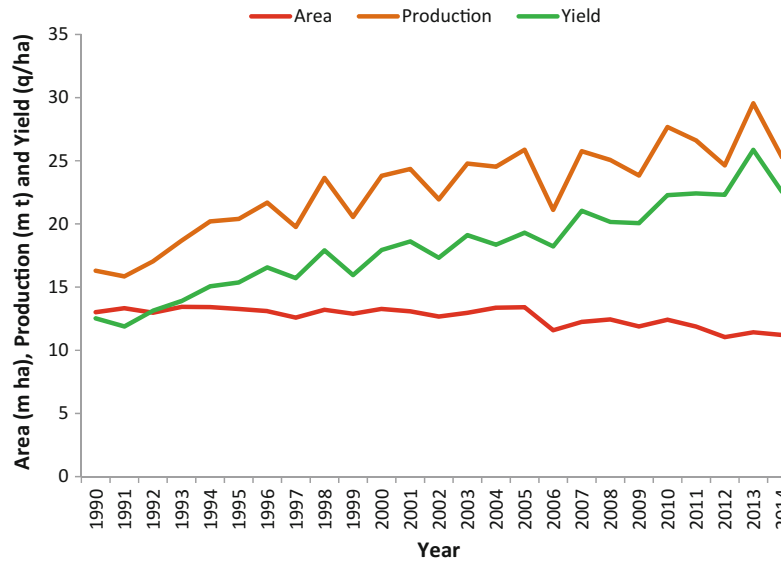


**Fig. 2.1** Yearly average variation for area, production and yield of peanut in Asia during the period from 1990 to 2014

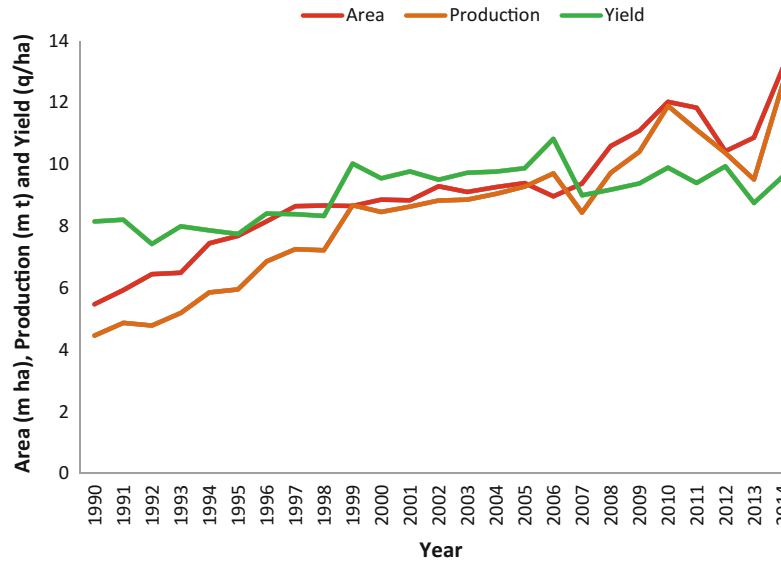


**Fig. 2.2** Yearly average variation for area, production and yield of peanut in Africa during the period from 1990 to 2014

improved seed, fertilizers, and pesticides resulting in low yields. One major constraint is the lack of improved varieties suited to different agroecological zones and availability of effective seed dispersal systems. Due to non-availability of seeds or efficient dissemination on information of new varieties, the majority of smallholder farmers still grow traditional landraces/outdated varieties that are adapted to local environments but have low genetic yield potentials and are, in many cases, susceptible to drought, pests, and diseases. Lack of efficient storage structures and low storability of peanut seeds under ambient conditions is another important constraint. Seed being a costly resource, many smallholder farmers tend to retain seeds from previous harvests for use in subsequent season. However, poor storage conditions and low use of seed-treatment chemicals further reduce the quality of the seed. The private seed sector

companies are hesitant to invest in peanut seed multiplication and distribution because the crop has high seed requirement, low multiplication factor, the bulky nature of the seed and low storability under normal conditions. Aflatoxin contamination of peanut by the fungi *Aspergillus flavus* and *A. parasiticus* is an important constraint affecting the quality of peanut in most producing countries in Africa and Asia. Aflatoxin contaminated peanuts are potent carcinogens for both humans and livestock and many importing countries have placed strict restrictions on acceptable levels of aflatoxin in peanut. Besides, the policy regulations restricting movement of seeds and marketable peanut products have hampered the development of the peanut sector, in many producing countries and especially in Africa.

### 2.1.4 Trading of Peanut

There is a great demand for peanut and peanut-based products in the International market, especially for confectionary types. In 2010–2013 an estimated 42 MMT (million metric tons) of peanuts were utilized annually in the world which was an increase of approximately 134% from 18 MMT in the 1970s (Fletcher and Shi 2014). Export or import to any country is guided by regulations and requires strict adherence to consistency in supply and in quality of the produce. Increasing the exports quantity is the primary objective as it earns foreign exchange for the country and also benefits the supply chain comprising of producers, processors, and traders. International trade of peanut is primarily in the form of pods (in-shell), shelled (kernels), meal (cake) and as oil. Over the past decades trade volumes have increased substantially for confectionary peanut but fallen for peanut oil and meal due to availability of cheaper substitutes such as soybean and raising concerns over aflatoxin contamination. Besides the higher nutritional quality of peanuts for food purpose, development of novel peanut-based food products and initiatives by the American Peanut Council for export promotion and education has led to improvements in existing peanut markets and developing new peanut-based markets for food use. In 2010–2013, about 41% of the world annual peanut production was crushed for domestic use, 45% for domestic food use and the remaining 14% was exported, used for feed or lost. Food use of peanut has increased by about 265% and crushing use increased by about 75% since the 1970s (Fletcher and Shi 2014).

#### 2.1.4.1 Peanut Oil

International trade of edible oil has declined over time as the major producers of the crop *viz*. China, India, and USA consume substantial amounts of edible oil in their domestic markets leaving very little surplus for export. The export trade of oil in the developing countries is concentrated mainly in Senegal and Sudan accounting for one-third of global exports. Among the developed countries, USA with 13% of world exports is the only significant exporter of peanut oil (Freeman et al. 1999). In 2008–2009, the USA, Argentina, Sudan, Senegal, and Brazil took about 71% share of the global export and the European Union (EU), Canada and Japan 78% share of global import in international trade of peanut (Rathnakumar et al. 2015).

#### 2.1.4.2 Peanut Kernels

Trading of shelled peanut has increased in recent times. Developing countries accounted for much of this increase. However, the utilization pattern varied across the regions. Fast-growing Asian economies such as China are devoting more peanuts to consumption due to rising per capita incomes and urbanization. During 2013, the global import of shelled peanut was 1.68 m tons estimated at US$ 2467 million while the global export stood at 1.67 m tons estimated at US$ 2195 million. India (388 k tons), Argentina (190 k tons), USA (177 k tons) and China [China (176 k tons), China mainland (176 k tons)] were the top five exporters of shelled peanuts during 2005–2013 while EU (448–497 k tons) was the top importer of shelled peanut followed by Netherlands (288 k tons) in the same period (FAOSTAT 2015).

#### 2.1.4.3 Peanut Meal

Peanut meal is an important source of protein for livestock. The production and trade of peanut meal is directly influenced by demands for peanut oil, competing prices between other oilseed meals and cereal-based-substitutes and the existing tariff barriers. During the period, 1979–1981 to 1994–1996, world utilization of peanut meal increased by 45% with most of the increase coming from Asia. The demand was more pronounced in Thailand, Indonesia, and some other rapidly developing Asian countries due to increased consumption of meat and livestock rearing. In the developed countries consumption declined by 60% during the same period, because of developments in the European market, which show the share of European Community in global utilization of peanut meal fall from 22% in 1979–1981 to 5% in 1994–1996. This decline was mainly due to increase in peanut meal prices, presence of aflatoxins in imported products beyond the permissible limit and its subsequent substitution by cheaper feed alternatives such as soybean meal.

The trading of peanuts in the International market requires strict adherence to aflatoxin content levels and is recognized as the primary non-tariff trade barrier for export of peanut by the developing countries such as Asia and Africa. The magnitude of losses incurred due to trading of aflatoxin contaminated peanuts is not known, but they have serious economic implications in terms of visible and invisible costs both at national and international level. The permissible limit varies among the countries—35 ppb (total) by Malaysia; 30 (ppb) total by India, Indonesia and Brazil; 20 ppb (total) by the USA, Kenya, and the Philippines; 15 ppb (total) by Canada, UAE and Australia; 10 ppb ($B_1$) by Japan, Korea, Taiwan and Singapore; 10 ppb (total) by Egypt and Vietnam; 5 ppb ($B_1$) by the Russian Federation and Turkey. However, the EU countries have set a very stringent maximum permissible limit of 2 ppb for $B_1$ and 4 ppb for total aflatoxins in peanuts. This new trade regulations led to decline in imports of peanut meal to EU countries from 0.91 Mt in 1979–1980 to 0.43 Mt in 1989–1990 (Bhat 1991). Similarly, the export of peanut meal from India declined from 550 t valued at US$ 42.5 million in 1977–1978 to 265 t valued at US$ 32.5 million in 1985–1986, mainly because of aflatoxins (Bhat and Rao 1990).

## 2.2 Nutritional Value

### 2.2.1 Kernels, Meals and Haulms

Peanut seeds (kernels), the most important product of peanut are a rich source of nutrition and provide several health benefits. The kernels contain 40–55% oil, 20–35% protein and 10–20% carbohydrate. They provide 567 kcal of energy from 100 g of kernels (Jambunathan 1991). The peanut oil contains seven fatty acids of which palmitic (7–12%), oleic (40–50%) and linoleic (25–35%) together account for approximately 90% of total fatty acids. High oleic lines containing >80% oleic acid are also available. Also, the seeds are good source of minerals like calcium, phosphorus, iron, and zinc; vitamins like E, and the B-complex groups of thiamin, pantothenic acid, riboflavin, foliates and niacin; antioxidants like *p*-coumaric acid and resveratrol; and biologically active polyphenols, flavonoids and isoflavones. Peanut meal obtained after oil extraction is a high protein rich feed for livestock and poultry. The primary constituents are crude protein (45.6%), sugar (32.50%), fat (2.5%), fiber (8.3%) and ash (5.0%). It is also a rich source of amino acids- lysine, methionine, cysteine, threonine and arginine, and minerals such as calcium, phosphorus, sodium, and potassium. The metabolizable energy of peanut meal is 2664 kcal/kg (Batal et al. 2005). Peanut haulms (the above ground vegetative part) is a good source of nutritious fodder for livestock, and contains protein (8–15%), lipids (1–3%), minerals (9–17%), crude fiber (22–38%) and carbohydrates (38–45%). It is used as cattle feed in fresh or dried stage, or by preparing hay or silage. Nutrient digestibility in the case of peanut haulms is around 53% and that of crude protein

is 88% when fed to animals (Nagaraj 1988). Haulms are capable of releasing energy up to 2.337 cal/kg of dry matter.

### 2.2.2 Food Products

Several value-added products developed from peanut are available around the world such as peanut flour, roasted and boiled peanut (in-shell/kernel), peanut butter, peanut candy etc. Local delicacies have also been developed for localized consumption purpose. For example, in many parts of Western Africa and Sudan partially defatted or full-fat peanut is a local delicacy and the most common form of utilization. Partially defatted peanut paste is produced after the oil has been extracted, and is used for making *kuli-kuli* in Nigeria and *coura-coura* in Burkina Faso. The full-fat peanut paste is a common food ingredient in Western Africa, Sudan and Southern Africa (Freeman et al. 1999).

### 2.2.3 Consumption Pattern

There is a visible divergence in the consumption pattern of peanut both in the developed and developing countries. Most of the peanut produced in the developing countries is crushed for extraction of oil to meet the domestic consumption needs, while in the developed countries such as USA it is mainly consumed as a food source. Over the years, even in developing countries, the trend has shifted more towards food source with increasing international market demands for confectionary grade peanuts and the availability of other cheaper alternative oils. In Africa, peanuts are consumed as roasted, boiled or raw and as peanut paste. In Argentina and Brazil, large quantities of confectionary peanut are consumed as roasted nuts or in packaged form as snack foods such as peanut candy. In USA, peanut consumption is mainly in the form of peanut butter, packaged snack nuts (salted, unsalted, flavored and honey-roasted) and peanut candies.

Even among countries, diversity exists in terms of regional preferences. For example, the food consumption of peanut dominates in North America and the oil consumption dominates in the South. In East and West Africa, both food and oil uses dominate while in South Africa the food use of peanut is dominant. In Southeast Asia, the food use dominates while in Southwest Asia, which is dominated by India, oil use is more important over food use (Rathnakumar et al. 2015).

### 2.2.4 RUTF and Food Supplements

Malnutrition is one of the most serious issues threatening the global community and especially in the developing countries. Globally it is estimated that about 20 million children suffer from severe acute malnutrition (SAM) of which about 8.1 million children are from India. One way of combating malnutrition issues is to provide the affected individuals with essential nutrients, minerals and vitamins in an easily available and ready to use form. Therapeutic foods are nutritionally enhanced food products, supplied in emergency situations for the treatment of SAM symptoms. Peanut is one of the important constituent of such product due to its balanced nutrient composition. Ready to use therapeutic food or RUTF originally referred to a nutrient dense and energy-dense peanut-based paste designed specifically for the treatment of SAM in young children. There are different types of RUTF currently available in the market, among which the 'Plumpy nut' patented product by Nutriset is widely recommended by UNICEF. This nutritional paste (peanuts, powdered milk, vegetable oil, sugar, vitamin and mineral mix) contains the right mix of nutrients to treat a child with SAM, and in a form that is easy to consume and safe. RUTFs provide 520–550 kcal/100 g (Kapil 2009; Dubey and Bhattacharya 2011).

Another product line Ready to use Supplementary Food (RUSF) is produced and marketed by Nutriset. RUSFs are foods that are fortified with micronutrients as a remedy for malnutrition and can be consumed without cooking or the addition of water. This product aims to tackle malnutrition at early stages (moderate acute

malnutrition, or in prevention of acute malnutrition or chronic malnutrition) and are used in addition to breastfeeding (for young children above 6 months of age) and traditional complementary food. The RUSF product line includes Plumpy doz, Supplementary Plumpy, QBMIX, and Delphia infant milk (Latham et al. 2011).

### 2.2.5 Functional Food Use of Peanuts

Peanuts are a good source of wide range of nutrients and bioactive compounds with health benefits. Most of these compounds are either present in the skin, the extracted oils and the kernels. Even the methanolic extracts from peanut hulls were reported to have strong antioxidant activity (Duh and Yen 1995) and ability of scavenging free radical and reactive oxygen species (Yen and Duh 1994). They are touted as functional foods due to the presence of numerous functional components like Coenzyme Q10. These bioactive components are widely recognized for their disease preventative properties. Some of the bioactive compounds such as tocopherols, tocotrienols, flavonoids and resveratrol function as antioxidants while others promote longevity.

#### 2.2.5.1 Tocopherols

Tocopherols (TCP) are a class of organic chemical compounds having vitamin E activity. Peanut oils are a good source of α and γ- tocopherols with contents varying from 50–373 ppm and 90–390 ppm respectively (Firestone 1999). The diversity depends on the origin (Sanders et al. 1992), variety, maturity and the growing conditions. Higher tocopherol content was consistently reported from US developed peanuts as compared to those produced in China or Argentina. Under same growing conditions runner varieties have higher levels of α-, γ- and δ- tocopherols than the Spanish varieties.

#### 2.2.5.2 Resveratrol

Resveratrol (3, 5, 4'-trihydroxystilbene) is a naturally occurring stilbene phytoalexin polyphenol. It is naturally produced by several plants in response to injury, stress, infection, or ultraviolet (UV) radiation (Jeandet et al. 2012). Resveratrols is reported to play positive roles in reducing cancer risks, heart diseases, tumor and inflammation (Arya et al. 2016). Peanuts are excellent sources of resveratrol with the southern style boiled peanuts having the most abundant, even more than that found in red wine and red grape juice on a part per million basis (Sanders et al. 2000) followed by peanut butter (Ibern-Gomez et al. 2000). All parts of the peanut contain resveratrol from the roots to the skins and even the shell (Francisco and Resurreccion 2008).

#### 2.2.5.3 Phytosterols

Phytosterols or plant sterols, a naturally occurring compound found in plant cell membranes are minor components of all vegetable oils and constitute major portion of the unsaponifiable fraction of the oil. Peanut oil contains 900–3000 ppm total phytosterols of which β-sitosterol (>80%), campesterol (10%) and stigmasterol (<5%) together constitute 95% (Firestone 1999). Phytosterols due to their structural similarity with cholesterol block the absorption of cholesterol in the digestive system thereby reducing the risk of cardiovascular diseases. People who consume small amounts of peanut daily were found to have lesser instance of heart-related diseases (Awad et al. 2000). Emerging evidence has shown that they also reduce inflammation and reduce the growth of various cancers (Woyengo et al. 2009).

#### 2.2.5.4 Arginine

Arginine is an amino acid that plays an important role in strengthening the body's immune system, regulating hormone and blood sugar levels and promoting male fertility. It is considered semi-essential because, although the body can manufacture its own supply, dietary supplementation may become essential under certain situations of severe injury and illness. Peanuts have the highest level of arginine among foods (USDA SR-21). Arginine is the precursor to nitric oxide which helps to keep the arteries relaxed, improve

the blood flow and healing time in tissues in the body (Moncada and Higgs 1993).

#### 2.2.5.5 Phenolic Acids and Flavonoids

Peanut and peanut skins are a good source of phenolic compounds (Francisco and Resurreccion 2008) especially p-coumaric acid. Peanut skins are often added to processed foods such as peanut paste and peanut butter to improve shelf-life, antioxidant capacity, and nutritional quality. Phenolic acids have been shown to play a protective role against oxidative damage diseases like coronary heart disease, stroke, and various cancers. It was further reported that roasted peanuts have phenolic acid levels comparable to those found in green tea and red wine, and more than those in berries when the skin is not removed (Francisco and Resurreccion 2008).

Flavonoids are a group of secondary biochemicals that mostly function in plant defense systems. They act as natural pesticide, some provide potent odors or bitter flavors as a defense system, while others are antimicrobial in nature. They are present throughout the peanut plant and are responsible for color, taste, and protection of vitamins, enzymes and fat oxidation. A high intake of flavonoids reduces chances of heart-related diseases and various types of cancer by diverse mechanisms which are still being researched.

## 2.3 Taxonomic Classification

### 2.3.1 Plant and Floral Biology

The cultivated peanut is an allotetraploid ($2n = 4x = 40$) and is believed to have originated from a cross involving the diploid species *A. duranensis* and *A. ipaensis* (Kochert et al. 1996; Seijo 2004, 2007). The main stem of the plant is either upright or prostrate (12–65 cm in length) and develops from a terminal bud of the epicotyl, while two cotyledonary laterals (prostrate, runner type or upright) grow on opposite sides. The stem usually bears tetrafoliate leaves, with leaflets on the main stem differing in shape and size from those on lateral branches.

Peanut flowers are typically papilionaceous and zygomorphic and represented either by a solitary flower (simple inflorescence) or by a raceme containing two to five flowers (compound inflorescence) in the axils of the cataphylls. The flowers are borne aerially but pod development takes place below the ground due to geotropic movement of the gynophores (pegs). Flowering in peanut is sensitive to light, temperature and relative humidity. Temperatures between 22 and 33 °C and soil moisture of 40% are ideal for flowering, while light intensity >45% of full sunlight helps in optimum floral development. Under normal conditions, flowers open at sunrise, but low temperature can delay the opening. Anther dehiscence can take place 7–8 h before flower opens in some varieties whereas in others they may not do so even at flower opening (Bolhuis et al. 1965). The stigma becomes receptive about 24 h prior to anthesis and its receptivity can persist for about 12 h after anthesis. Self-pollination takes place within the closed keel of the flower. About 40% of the flowers fail to begin pod development and another 40% abort before pod development.

### 2.3.2 Center of Origin and Distribution

The exact center of origin of peanut is unclear but it is believed to be somewhere in the region of eastern foothills of the Andes (southern Bolivia to northwestern Argentina) because of the primitive characters (pod beak, pod shape, pod reticulation etc.) associated with germplasms from the region (Krapovikas 1969; Gregory et al. 1980). It is naturally restricted to Argentina, Bolivia, Brazil, Paraguay, and Uruguay in South America. The greatest genetic diversity in *Arachis* was reported in South America with six recognized gene centers for cultivated peanut in South America—(i) the Guarani region, (ii) Goias and Minas Gerais (Brazil), (iii) Rondonia and northwest Mato Grosso (Brazil), (iv) the eastern foothills of the Andes in Bolivia, (v) Peru, and (vi) Northeastern Brazil. A seventh center Ecuador was added to the group following

identification of distinct group of landraces referred as var. *aequatoriana* (Krapovickas and Gregory 1994, 2007).

The domestication of peanut probably happened in the valleys of the Parana and the Paraguay river systems in the Gran Chaco area of South America (Hammons 1994). Remnant single-seeded peanut shells recovered from archeological excavations in coastal Peru dating back to 800 BC evidenced the cultivation of peanut. From South America, the peanut spread to other parts of the world. The 'Virginia variety' was taken from the Antilles to Mexico around 1500 and then quickly introduced into West Africa. Subsequently, it was introduced into North America in the 17th century. Portuguese explorers in the late 15th century carried 2-seeded 'Spanish' peanut varieties from South America (Brazil) to Africa, where it got mixed with the 'Virginia' types and produced a great diversity of African land races. The Spaniards in the early 16th century took 3-seeded Peruvian types (including hirsuta types) to Philippines and then to southeastern China where it was referred to as 'foreign beans' (Nigam 2015). From there it spread throughout China and to Japan as 'Chinese beans'. The 'Valencia types' were taken from Cordoba, Argentina around 1900 and introduced into Spain and subsequently to USA from Valencia during 1910 (Rathnakumar et al. 2013).

### 2.3.3 Classification

The genus *Arachis* based on morphology, geographical distribution and cross compatibility has been divided into nine taxonomic sections and comprises of 80 described species (Krapovickas and Gregory 1994, 2007; Valls and Simpson 2005), which includes both diploids and tetraploids belonging to either annual or perennial type. Among them *Arachis hypogaea* L. is the only cultivated species. It is a tetraploid (amphidiploid or allotetraploid) with a chromosome number $2n = 4x = 40$. Besides, *A. hypogaea*, two other species *A. villosulicarpa* (cultivated in northwestern Brazil) and *A. stenosperma* (cultivated in central and southwestern Brazil) are grown for their seeds.

The cultivated peanut is divided into two subspecies, sub sp. '*hypogaea*' and subsb. '*fastigiata*' based on the branching pattern and the distribution of vegetative and reproductive axes. The former subspecies is characterized by the absence of reproductive axes (flowers) on the main stem and the presence of alternate pairs of vegetative and reproductive axes on the cotyledonary laterals and $n + 1$ lateral branches (called alternate branching pattern). The latter is characterized by the presence of reproductive axes on the main stem and the presence of reproductive axes on successive nodes of lateral branches (called sequential branching). The subsp *fastigiata* is comprised of four botanical varieties, var. *fastigiata*, var. *vulgaris*, var. *peruviana*, and var. *aequatoriana*, while subsp *hypogaea* is divided into two varieties, var. *hypogaea* and var. *hirsuta* based on inflorescence, pod and seed characters.

### 2.3.4 Market Types

Based on popularity and market uses four types of peanut has been defined in the United States: Spanish, Runner, Virginia, and Valencia. The large-seeded Virginia types are the most widely cultivated peanut in the Virginia-North Carolina area; the runner market type is grown predominantly in the Southeast and Southwest America, and the Spanish types are grown in Texas and Oklahoma. The Valencia market types are mostly produced in New Mexico for the in-shell market (Holbrook and Stalker 2003). Depending on differences in flavor, oil content and quality, size and shape of pods and kernels certain types are preferred over others; but for most cases the different types are interchangeable. In the US most peanuts marketed in the shell are of the Virginia type, along with some Valencia selected for large size and attractive appearance of the shell. Valencia peanuts are very sweet in taste and are also excellent for consumption as boiled peanuts. The Spanish types are mostly used for

making peanut candy, salted nuts, and peanut butter. The runner types are mostly preferred for making peanut butter.

## 2.4 Peanut Research

### 2.4.1 Breeding New Varieties

Hybridization between selected parents, selection using phenotyping and advancing the generations, followed by yield trials have led to development and release of several varieties suitable to varying production environments and meet the needs of the producers, consumers, and industry (Janila and Nigam 2013). Among the different traits, breeding for high yield is the most important yield component for determining performance of new varieties, although kernel yield and oil yield are considered under special circumstances such as for developing high oil lines. In India 194 peanut varieties have been released by 2012 which have contributed to increased yield. However, most of the yield improvements came through increase in number of pods and improvements in pod and seed size (Reddy 1988; Ratnakumar et al. 2010, 2013). Varieties such as Vijetha, Girnar-3, GPBD-5, ICGV 00350, RARS-T-1, GJG-31, GJG-9 were released following their superior yield performance in national trials (Ratnakumar et al. 2013). Starting from 1976, ICRISAT has developed and released 179 peanut varieties across 38 countries globally. In China, yields of 9 t/ha were obtained from improved cultivars when grown under favorable conditions (Yu 2011). Along with yield, the length of the growing period (LGP) and resistance to pathogens and pests are important to enable adaptation of peanut to new regions or special cropping systems.

Among the fungal diseases, early leaf spot (causal agent *Cercospora arachidicola*), late leaf spot (*Phaeoisariopsis personata*) and rust (*Puccinia arachidis* Spegazzini) are important foliar pathogens and are the focus of most peanut breeders across the world. Varieties with resistance to foliar fungal diseases were reported (Singh et al. 1997). Very high levels of resistance to foliar fungal diseases occurs in related wild species of peanut but has limited utility as a consequence of undesirable genetic linkage between resistance and low yield, late maturity, low shelling outturn, heavy pod reticulation, bitter kernels etc. (Liao 2014). Bacterial wilt (BW) caused by *Ralstonia solanacearum* is a major production constraint of peanut in China, Indonesia, Vietnam, and Uganda. Breeding efforts for BW resistance concentrated on screening and identifying BW resistant lines (Singh et al. 1997; Hong et al. 1999). Among the viruses groundnut rosette disease (GRD) in Africa, peanut bud necrosis disease (PBND) in India, tomato spotted wilt virus (TSWV) in East and South east Asia, peanut stem necrosis disease (PSND) in some areas in southern India, and peanut clump virus disease (PCVD) in West Africa are major breeding targets worldwide. Breeding for virus resistance has achieved significant progress with the identification of resistant and tolerant lines both among cultivated and wild *Arachis* sp. (Upadhyaya et al. 2011; Nigam 2015).

Both, physiological trait-based and empirical selection approaches are used for improving drought tolerance in peanut. Breeding heat tolerant genotypes has become a priority with changing climatic conditions and increase in temperature. Nutritional quality aspect of peanut is gaining importance worldwide with the development of high-end tools for quality assessment and their requirement in different products. For example, in confectionary peanut, the quality attributes targeted include high sugar, high protein, low oil, attractive seed size and shape, pink or tan seed color, ease of blanching and high oleic/linoleic ratio, while for developing RUTF and food supplement based products, peanuts with high protein, minerals and vitamins

are preferred. For edible oil and biofuel purpose, varieties with high oil content and specific fatty acid profiles are desired.

### 2.4.2 Genetics of Important Agronomic Traits

A thorough knowledge of nature of inheritance, interaction with the environment, the nature of gene action and the number of alleles/genes involved in governing agronomically important traits is key to target their improvement. Most agronomic traits in peanut are inherited quantitatively and are highly influenced by genotype × environment interactions. The genetics of several important target traits in peanut have been studied and this information is well documented (Reddy and Murthy 1996; Nigam 2015). In peanut, pod yield is the most important and complex trait and it is associated with over 40 other traits (Murthy and Reddy 1993). Genetic studies have identified both additive and non-additive components of genetic variances to be important for yield and related traits. Significant cytoplasmic influence on yield and related characters was also observed (Dwivedi et al. 1989). Oil content in peanut is controlled by both additive and non-additive components of gene action. The low levels of genetic variability were a major hindrance in breeding for high oil content in peanut seeds. Identification of high oil lines both among cultivated and wild *Arachis* species has accelerated breeding efforts to develop cultivars with oil content higher than 55%. Iodine value, an indicator of oil quality and stability has been reported to be governed predominantly by additive gene action (Basu et al. 1988). The high oleic trait in peanut is controlled by two recessive genes located on the A and B genomes (Knauft et al. 1993).

Genetic studies for drought tolerance are mainly restricted to its contributing surrogate traits as it is very difficult to measure it under field conditions. Sufficient variability for physiological traits such as specific leaf area (SLA), soil water extraction ability, water use efficiency and harvest index (HI) was observed among tolerant and susceptible genotypes. It was reported that both additive and additive × additive gene effects for SLA and HI and additive gene effects for $\Delta\ ^{13}C$ (carbon isotope discrimination) are the major genetic factors (Nigam et al. 2001).

Studies on genetics for rust resistance in cultivated peanut has revealed that it is governed by two or more recessive genes interacting in various ways (Nigam 2015). However, the resistance in wild *Arachis* species is controlled by dominant genes (Singh et al. 1984). Resistance to early and late leaf spot has been reported to be independently controlled by two or more major genes (Tiwari et al. 1984) and several minor genes predominantly with additive effects (Anderson et al. 1986). In the wild species, resistance to ELS and LLS was reported to be independently inherited (Nigam 2015). Aflatoxin contamination is a major problem in large and extra-large kernelled peanut genotypes and those exposed to drought stress. Three levels of resistance mechanism were identified—preharvest resistance, seed coat resistance (in vitro seed colonization) and cotyledon resistance (aflatoxin production). For *A. flavus* infection a pair of major genes with additive value of 0.38 and a pair of minor genes with additive value of 0.12 was reported in literature (Zhou et al. 1999; Zhou and Liang 2002). Seed coat resistance has been reported to be controlled by predominant additive genes and maternal genotype (Rao et al. 1989). Resistance to PBND has been reported to be governed by three factors (Pensuk et al. 2004). For TSWV significant general combining ability (GCA), specific combining ability (SCA) and transgressive segregation was reported but the genetic mechanism of resistance is yet to be elucidated. In the case of GRD (effective against GRV and its SatRNA) resistance in cultivated types is

reported to be governed by two independent recessive genes which are effective against both chlorotic and green rosette (Olorunju et al. 1992).

### 2.4.3 Genomic Tools and Genome Sequence

Use of genomics based approaches for improvement of economically important target traits in peanut has been challenging due to its inherent genetic architecture. Narrow genetic base of the primary gene pool, tetraploid nature of the cultivated peanut and cultivation of limited genotypes in the process of domestication has resulted in diminishing genetic resources and low variability for several traits. Presence of quantifiable variability is a must to identify linked molecular markers and/or quantitative trait loci (QTLs) for marker-assisted breeding for crop improvement. Among all genomic tools, molecular markers have proved to be the most useful in characterizing and harnessing available genetic variations. The early generation markers were basically used for conducting genetic diversity studies (Bravo et al. 2006), in limited cases for construction of genetic maps (Garcia et al. 2005; Leal-Bertioli et al. 2009) and identification of associated genes/QTLs (Herselman et al. 2004). The development of more efficient marker systems such as Simple Sequence Repeat (SSR), Single Nucleotide Polymorphism (SNP) etc. led to identification of closely linked markers for several target traits such as resistance to nematode, ELS and LLS, rust, high oleic acid, drought tolerance and their utilization in breeding programs worldwide (Pandey et al. 2014). The different marker systems in peanut and their utilization in trait breeding is reviewed in Janila et al. (2016).

Although genetic mapping studies of peanut started in the late 20th century (Halward et al. 1993), the first report of genetic map of cultivated peanut was published in 2009 (Varshney et al. 2009). Since then, efforts have been directed towards refining the genetic map using mapping populations (Khedikar et al. 2010) or through construction of composite linkage maps (Hong et al. 2010) and integrated maps (Qin et al. 2012). The first international reference consensus map for tetraploid peanut was constructed by Gautami et al. (2012) based on data obtained from 11 populations. The map had 897 marker loci (895 SSR loci and two cleaved amplified polymorphic sequences (CAPS)) distributed on 20 linkage groups and spanning a map distance of 3863.6 cM with an average map density of 4.4 cM. Considering the huge potential offered by SNPs in marker trait association studies, efforts were also made to develop SNP based linkage maps in peanut. The first SNP marker based genetic map was developed for the AA genome of peanut (Nagy et al. 2012). This was followed by the development of an SNP based linkage map for the cultivated peanut. The linkage map was constructed using 1685 marker loci (1621 SNPs and 64 SSRs) spanning a distance of 1446.7 cM (Zhou et al. 2014).

With the advent of Next-Generation Sequencing (NGS) technology platforms, sequencing the peanut genome has now become a distinct possibility. NGS technologies offer faster sequence data generation and informatics tools to manage and analyze NGS data (Varshney and May 2012) in a relatively short time. To sequence the peanut genome, the Peanut Genome Consortium (PGC) was formed for the tetraploid cultivar "Tifrunner".[2] Very recently, the genome sequences of *A. duranensis* and *A. ipaensis*, the diploid ancestors of cultivated peanut was completed (Bertioli et al. 2016). The sequence information will be useful to identify candidate disease resistance genes, to develop molecular markers, to guide tetraploid transcript assemblies and to detect genetic exchange between cultivated peanut's subgenomes.

### 2.4.4 Aflatoxin and Allergens

Two quality deterrents- aflatoxin contamination and allergens play a significant role in determining the industry and consumer base and the marketability of the produce. Contamination of

[2] http://www.peanutbioscience.com/peanutgenomeproject.html.

peanuts by aflatoxin is a global issue forcing many countries to have strict restrictions with regards to aflatoxin content in the produce. Aflatoxins are secondary metabolites produced by colonization of peanut kernels by *Aspergillus flavus* (Link) and *Aspergillus parasiticus* (Speare) and are considered among the most potent carcinogenic mycotoxins in nature. Peanuts are susceptible to *Aspergillus* infection and aflatoxin contamination and the infection can occur either in the field, during post-harvest drying or during curing and storage. Three types of resistance to *Aspergillus* infection and aflatoxin production have been reported in peanut operating at pods (preharvest resistance), seed coat (in vitro seed colonization (IVSC)) and cotyledon levels (aflatoxin production) (Utomo et al. 1990; Nigam 2015). Screening techniques for evaluating resistance of genotypes and advanced generation populations under controlled and field conditions for resistance to seed infection by the fungi and resistance to aflatoxin formation were developed and resistant sources identified. Resistant sources such as ICG 1122, ICG 1326, ICG 3263, ICG 3336 for preharvest infection; PI 337394F, Ah 78223, Monir 240–30 for IVSC resistance; and ICG 10609, ICG 11682, ICG 9610 for aflatoxin production are available, but none of the genotypes are completely free from infection. A basic drawback in identifying resistant lines has been the inconsistency between in vitro resistance screening and field resistance testing (Anderson et al. 1995). Studies have reported a very low correlation (−0.07) between IVSC by *Aspergillus flavus* (IVSCAF) and seed infection in the field indicating independent resistance genetic mechanisms for both types of infection (Utomo et al. 1990; Upadhyaya et al. 1997). For example, in screening trials conducted in the US, it was found that genotypes reported to be resistant to IVSCAF or preharvest aflatoxin contamination performed similar to the susceptible cultivar Florunner in levels of aflatoxin contamination when subjected to an extended period of heat and drought stress (Anderson et al. 1995). Therefore the sampling procedures and screening methods including development of infector plot need to be further refined to improve uniformity of infection, characterization, and precision of estimation of infection and aflatoxin production in a genotype in a consistent manner (Nigam 2015). Functional genomic tools such as microarray technology, expressed sequence tags (ESTs) are being utilized to identify genes that are expressed or repressed under *Aspergillus* infection (Luo et al. 2005; Guo et al. 2008, 2011) and also those that influence aflatoxin contamination levels (Guo et al. 2008), but developing zero aflatoxin peanuts still remains a dream for peanut researchers across the world. Transgenic technologies involving silencing key genes that regulate aflatoxin biosynthetic pathways holds great promise in this regard.

Peanut proteins are regarded as a major source of allergens and ingestion of seeds is reported to be one of the most serious causes of fatal food-induced anaphylaxis (Yocum and Khan 1994). Peanut-induced anaphylaxis is not a major problem in the Asian and African countries but is more severe in the USA where 0.8% of children and 0.6% of adults are allergic to peanut protein (Sampson 2004; Nigam 2015). Thus, developing non allergenic peanut cultivars is a highly desirable objective among the scientific community. Studies on the nature of the allergen causing compounds have revealed the involvement of about 13 peanut allergens (www.allergen.org) of which Ara h1, Ara h2 and Ara h3 are classified as the major peanut allergens because they are generally recognized by more than 50% of peanut-allergic patients (Koppelman et al. 2001). Specifically, Ara h1 and Ara h2 are recognized by 70–90% of patients with peanut allergy (Burks et al. 1995; Clarke et al. 1998), and Ara h3 is recognized by serum IgE from approximately 44% to 54% of different patient populations with a history of peanut sensitivity (Rabjohn et al. 1999). Screening of genetically diverse peanut germplasms indicated that variability for Ara h1 ranged from 7 to 18.5%, Ara h2 from 5.9 to 13.2%, and Ara h3 from 21.8 to 38.5% of the total protein content of the seed (Koppelman et al. 2001; Kang et al. 2007).

Breeding for reducing or modifying the allergenic proteins through natural or induced

variations or complete elimination through bio-engineering tools such as gene deletion, gene silencing or reduced gene expression are being utilized in peanut improvement programs across the world. Very little natural variation exists in *A. hypogaea* for allergenecity. When the US mini core collection was evaluated for variation in allergen gene expression levels a 2-fold variation was observed in protein amounts for the three major allergens (Kang et al. 2007). Targeting one allergenic component such as Ara h1 for reduced/complete loss of functionality does not seem to have much effect as it is most often compensated for by the presence of other allergens. Molecular tools such as post-transcriptional gene silencing to knock out the production of allergenic protein (Ara h1 and Ara h2) are being employed to mitigate the allergen problem (Dodo et al. 2005, 2008). Recently, genome sequencing of the A-genome progenitor of peanut, *A. duranensis* revealed 21 candidate allergen-encoding genes of which 9 are already reported in cultivated peanut (Chen et al. 2016). Sequence information and functional characterization of these allergen-encoding genes will be useful to identify genetic or medical interventions to allergy mitigation.

### 2.4.5 Genetic Transformation of *Arachis*

In peanut, the first successful transgenic plant was achieved using the genotype independent method of biolistic/bombardment technique in 1993 (Ozias-Akins et al. 1993). Subsequently, different protocols were developed for transformation, selection and regeneration of transformants which either utilized the genotype independent biolistic approach of targeting embryonic tissues (Chu et al. 2013) or the *Agrobacterium* mediated transformation using shoot regeneration cultures (Sharma and Anjaiah 2000). Peanut tissues such as leaf sections, embryo axes, hypocotyls, cotyledonary nodes etc. have been targeted for *A. tumefaciens* transformation with different success rates depending on the culture conditions, cocultivation protocols and host-pathogen interactions (Holbrook et al. 2011).

Transgenic peanut expressing genes for traits such as resistance to virus, insect and fungus, drought tolerance and grain quality have been developed by different research groups particularly in India, China, and the United States. The first transgenic peanut harboring the herbicide resistance *bar* gene was developed in 1994 (Brar et al. 1994). Another transgenic peanut with tolerance to the herbicide paraquat was developed by the transfer of Bcl-xL gene (Chu et al. 2008). However, high levels of Bcl-xL gene expression was found to be deleterious for plant cells. In the case of virus resistance, transgenic peanuts resistant to TSWV, Peanut Stripe Virus (PStV), Tobacco Streak Virus (TSV) were developed either through transfer of key viral genes in sense or antisense direction (to silence expression of viral proteins) (Li et al. 1997; Yang et al. 1998) or by expressing viral genes in the transgenic plants (Higgins et al. 2004; Mehta et al. 2013). For conferring insect pest resistance, the *cry1EC* δ-endotoxin gene from *Bacillus thuringiensis* was transferred to cultivated peanut and showed effective protection against the larvae of tobacco cutworm (*Spodoptera litura*) (Tiwari et al. 2008). Resistance to the necrotropic fungus *Sclerotinia minor* and *Sclerotinia sclerotiorum*, fungus responsible for causing Sclerotinia blight in peanut, was achieved by transferring the oxalate oxidase gene from barley into three Virginia peanut cultivars (Livingstone et al. 2005). The transformed plants had significantly reduced lesion size when compared to their respective nontransformed control cultivars.

Drought-tolerant peanut plants were reported using *AtDREB1A*, a cis-acting factor that binds to dehydration responsive element (DRE) from *Arabdiopsis thaliana* under the control of a stress inducible promoter rd29A gene (Bhatnagar-Mathur et al. 2009) and isopentenil-transferase (*ipt*) gene isolated from *Agrobacterium tumefaciens*, under the control of SARK (a drought-inducible promoter from bean, *Phaseolus vulgaris*) (Qin et al. 2011). IPT is a key enzyme in the biosynthesis of cytokinins, a plant phytohoromone which plays important role

in root growth and development. The multigenic trait gene *AtNHX1* from *Arabidopsis* was transformed into peanut and the transformants had enhanced drought and salinity tolerance (Asif et al. 2011; Banjara et al. 2012). The gene *AtNHX1* is an Na +/H + antiporter in *A. thaliana* and its over expression increases the ability to sequester sodium into vacuoles, thereby reducing cytosol toxicity, favoring water uptake by root cells and improving tissue retention under stress conditions. Similarly, over expression of *AVP1* (Qin et al. 2013) and *mtlD* (Bhauso et al. 2014) gene in peanuts resulted in increased salt and salinity tolerance. *AVP1* encodes a H + pyrophosphatase with proton pump activity on vacuoles, while *mtlD* gene encodes the enzyme Mannitol 1-Phosphatase Dehydrogenase that converts mannitol 1-phosphate to mannitol which is accumulated in the transgenic tissue. Although significant progress has been made in transgenic peanuts development, to date no released peanut cultivars are transgenic. Most of the developed products are under evaluation at different containment levels: in vitro, greenhouse and field conditions.

## References


Anderson WF, Holbrook CC, Wilson DM, Matheron ME (1995) Evaluation of preharvest aflatoxin contamination in some potentially resistant peanut genotypes. Peanut Sci 22:29–32

Anderson WF, Wynne JC, Green CC, Beute MK (1986) Combining ability and heritability of resistance to aarly and late leafspot of peanut. Peanut Sci 13:10–14

Arya SS, Salve AR, Chauhan S (2016) Peanuts as functional food: a review. J Food Sci Technol 53 (1):31–41

Asif MA, Zafar Y, Iqbal J, Iqbal MM, Rashid U, Ali GM, Arif A (2011) Enhanced expression of AtNHX, in transgenic groundnut (*Arachis hypogaea* L.) improves salt and drought tolerence. Mol Biotechnol 49:250–256

Awad AB, Chan KC, Downie AC, Fink CS (2000) Peanuts as a source of β-sitosterol, a sterol with anticancer properties. Nutr Cancer 36(2):238–241

Banjara M, Zhu L, Shen G, Payton P, Zhang H (2012) Expression of an *Arabidopsis* sodium/proton antiporter gene (AtNHX1) in peanut to improve salt tolerance. Plant Biotechnol Rep. 6:59–67

Basu MS, Nagraj G, Reddy PS (1988) Genetics of oil and other major biochemical components in groundnut (*Arachis hypogaea* L.). Indian J Trop Agri 6:106–110

Batal A, Dale N, Café M (2005) Nutrient composition of peanut meal. J Appl Poult Res 14:254–257

Bell SL (2008) Peanuts and their classification under the HTSUS. http://www.cbp.gov

Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD et al (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet 48:438–446

Bhat RV (1991) Aflatoxins: successes and failures of three decades of research. In: Fungi and Mycotoxins in stored products, Proceedings of an International conference held at Bangkok, Thailand, 23–26 April 1991, p 80–85

Bhat RV, Rao RN (1990) Practical guide for prevention and control of aflatoxins in groundnut. National Institute of Nutrition, Hyderabad, India, p 16

Bhatnagar-Mathur P, Devi MJ, Vadez V, Sharma KK (2009) Differential antioxidative responses in transgenic peanut bear no relationship to their superior transpiration efficiency under drought stress. J Plant Physiol 166(11):1207–1217

Bhauso TD, Thankappan R, Kumar A, Mishra GP, Dobaria JR, Rajam M (2014) Over-expression of bacterial mtlD gene confers enhanced tolerance to salt-stress and water-deficit stress in transgenic peanut (*Arachis hypogaea* L.) through accumulation of mannitol. Aust J Crop Sci 8:413–421

Bolhuis GG, Frinking HD, Leeuwaugh I, Reno RG, Staritsky G (1965) Observations on the opening of flowers, dehiscence of anthers and growth of pollen tubes in *Arachis hypogaea*. Netherlands J Agri Sci 7:138–140

Brar GS, Cohen BA, Vick CL, Johnson GW (1994) Recovery of transgenic peanut (*Arachis hypogaea* L.) plants from elite cultivars utilizing ACCELL® technology. Plant J 5:745–753

Bravo JP, Hoshino AA, Angelici CMLCD, Lopes CR, Gimenes MA (2006) Transferability and use of microsatellite markers for the genetic analysis of the germplasm of some *Arachis* section species of the genus *Arachis*. Genet Mol Biol 29:516–524

Burks AW, Cockrell G, Stanley JS, Helm RM, Bannon GA (1995) Recombinant peanut allergen Ara h I expression and IgE binding in patients with peanut hypersensitivity. J Clin Invest 96:1715–1721

Chen X, Li H, Pandey MK, Yang Q, Wang X et al (2016) Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insight into geocarpy, oil biosynthesis, and allergens. PNAS 113(24):6785–6790

Chu Y, Deng XY, Faustinelli P, Ozias-Akins P (2008) Bcl-xL transformed peanut (*Arachis hypogaea* L.) exhibits paraquat tolerance. Plant Cell Rep 27:85–92

Chu Y, Bhattacharya A, Wu C, Knoll J, Ozias-Akins P (2013) Improvement of peanut (*Arachis hypogaea* L.) transformation efficiency and determination of

transgene copy number by relative quantitative real-time PCR. In Vitro Cell Dev Biol -Plant 49:266–275
Clarke MC, Kilburn SA, Hourihane JO, Dean KR, Warner JO, Dean TP (1998) Serological characteristics of peanut allergy. Clin Exp Allergy 28:1251–1257
Dodo HW, Konan KN, Viquez OM (2005) A genetic engineering strategy to eliminate peanut allergy. Curr Allergy Asthma Rep 5:63–73
Dodo HW, Konan KN, Chen FC, Egnin M, Viquez OM (2008) Alleviating peanut allergy using genetic engineering: the silencing of the immunodominant allergen Ara h 2 leads to its significant reduction and a decrease in peanut allergenicity. Plant Biotechnol J 6:135–145
Dubey A, Bhattacharya M (2011) Ready to use therapeutic food: a review. Ind J Pract Pediatrics 13:1–50
Duh PD, Yen GC (1995) Changes in antioxidant activity and components of methanolic extracts of peanut hulls irradiated with ultraviolet light. Food Chem 54:127–131
Dwivedi SL, Thendapani K, Nigam SN (1989) Heterosis and combining ability studies and relationship among fruit and seed characters in peanut. Peanut Sci 16:14–20
FAOSTAT (2015) faostat3.fao.org (last accessed on June 25, 2016)
Firestone D (1999) Physical and chemical characteristics of oils. AOCS Press, Champaign, IL, Fats and Waxes
Fletcher SM, Shi Z (2014) An overview of world peanut markets. In: Stalker T, Wilson RF (eds) Peanuts: genetics, processing and utilization. AOCS Press, Elsevier Inc, London, UK, pp 267–287
Francisco ML, Resurreccion AV (2008) Functional components in peanuts. Crit Rev Food Sci Nutr 8 (8):715–746
Freeman HA, Nigam SN, Kelley TG, Ntare BR, Subrahmanyam P, Boughton D (1999) The world groundnut economy: facts, trends and outlook. ICRISAT, Patancheru, India, p 52
Garcia GM, Stalker HT, Shroeder E, Lyerly JH, Kochert G (2005) A RAPD-based linkage map of peanut based on a backcross population between the two diplod species *Arachis stenosperma* and *A. cardenasii*. Peanut Sci 32:1–8
Gautami B, Foncčka D, Pandey MK, Moretzsohn MC, Sujay V, Qin H, Hong Y, Faye I, Chen X, Bhanu Prakash A, Shah TM, Gowda MVC, Nigam SN, Liang X, Hoisington DA, Guo A, Bertioli DJ, Rami Jean-Francois, Varshney RK (2012) An international reference consensus genetic map with 897 marker loci based on 11 mapping populations for tetraploid groundnut (*Arachis hypogaea* L.). PLoS ONE 7:1–11
Gregory WC, Krapovickas A, Gregory MP (1980) Structure variation, evolution, and classification in *Arachis*. In: Summerfield RJ, Bunting AH (eds) Advances in Legume Science. Royal Botanic Gardens, Kew, pp 469–481
Guo B, Chen X, Dang P, Scully BT, Liang X, Holbrook CC, Yu J (2008) Peanut gene expression profiling in developing seeds at different reproduction stages during *Aspergillus parasiticus* infection. BMC Dev Biol 8:12
Guo B, Fedorova ND, Chen X, Wan CH, Wang W, Nierman WC, Bhatnagar D (2011) Gene expression profiling and identification of resistance genes to *Aspergillus flavus* infection in peanut through EST and microarray strategies. Toxins (Basel) 3:737–753
Halward T, Stalker HT, Kochert G (1993) Development of an RFLP linkage map in diploid peanut species. Theor Appl Genet 87:379–384
Hammons RO (1994) The origin and history of Groundnut. In: Smartt J (ed) The groundnut crop. a scientific basis for improvement. Chapman & Hall, London, pp 24–42
Hathorn CS, Sanders TH (2012) Flavor and antioxidant capacity of peanut paste and peanut butter supplemented with peanut skins. J Food Sci 77:S407–S411
Herselman LR, Thwaites FM, Kimmins B, Curtois PJA, Merwe VD, Seal SE (2004) Identification and mapping of AFLP markers linked to peanut (*Arachis hypogaea* L.) resistance to the aphid vector of groundnut rosette disease. Theor Appl Genet 109:1426–1433
Higgins CM, Hall RM, Mitter N, Cruickshank A, Dietzgen RG (2004) Peanut stripe potyvirus resistance in peanut (*Arachis hypogaea* L.) plants carrying viral coat protein gene sequences. Transgenic Res 13:59–67
Holbrook CC, Stalker HT (2003) Peanut breeding and genetic resources. In: Janick J (ed) Plant breeding reviews. Wiley, p 297–356
Holbrook CC, Ozias-Akins P, Chu Y, Guo B (2011) Impact of molecular genetic research on peanut cultivar development. Agronomy 1:3–17
Hong NX, Mehan VK, Lieu NV, Yen NT (1999) Identification of groundnut genotypes resistant to bacterial wilt in Vietnam. Int J Pest Manag 45:239–243
Hong Y, Chen Y, Liang X, Liu H, Zhou G, Li S, Wen S, Holbrook CC, Guo B (2010) A SSR-based composite genetic linkage map for the cultivated peanut (*Arachis hypogaea* L.) genome. BMC Plant Biol 10:10–17
Ibern-Gomez M, Sonia Roig-Perez, Lamuela-Raventos RM, Torre-Boronat MC (2000) Resveratrol and piceid levels in natural and blended peanut butters. J Agric Food Chem 48(12):6352–6354
Jambunathan R (1991) Groundnut quality characteristics. In: Uses of tropical grain legumes: proceedings of a consultant meeting, ICRISAT, Patancheru, 27–30 March 1989, p 265–275
Janila P, Nigam SN (2013) Phenotyping for groundnut (*Arachis hypogaea* L.) improvement. In: Panguluri SN, Kumar AA (eds) Phenotyping for plant breeding: applications of Phenotyping methods for crop improvement. Springer, New York, pp 129–167
Janila P, Variath MT, Pandey MK, Desmae H, Motagi BN, Okori P, Manohar SS, Rathnakumar AL, Radhakrishnan T, Liao B, Varshney RK (2016) Genomic tools in groundnut breeding program: status and perspectives. Front Plant Sci 7:289

Jeandet P, Delaunois B, Aziz A, Donnez D, Vasserot Y, Cordelier S, Courot E (2012) Metabolic engineering of yeast and plants for the production of the biologically active hydroxystilbene, resveratrol. J Biomed 2012:579089. doi:10.1155/2012/579089

Kang I, Gallo M, Tillman BL (2007) Distribution of allergen composition in peanut (*Arachis hypogaea* L.) and wild progenitor (*Arachis*) species. Crop Sci 47:997–1003

Kapil U (2009) Ready to use therapeutic food in the management of severe acute malnutrition in India. Ind Pediatr 46:381–382

Khedikar YP, Gowda MVC, Sarvamangala C, Patgar KV, Upadhyaya HD, Varshney RK (2010) A QTL study on late leaf spot and rust revealed one major QTL for molecular breeding for rust resistance in groundnut (Arachis hypogaea L.). Theor Appl Genet 121:971–984

Knauft DA, Moore KM, Gorbet DW (1993) Further studies on the inheritance of fatty acid composition in peanut. Peanut Sci 20:74–76

Kochert G, Stalker HT, Gimenes M, Galgaro L, Lopes CR, Moore K (1996) RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). Am J Bot 83:1282–1291

Koppelman SJ, Vlooswijk RA, Knippels LM, Hessing M, Knol EF, van Reijsen FC, Bruijnzeel-Koomen CA (2001) Quantification of major peanut allergens Ara h 1 and Ara h 2 in the peanut varieties Runner, Spanish, Virginia, and Valencia, bred in different parts of the world. Allergy 56:132–137

Krapovickas A (1969) The origin, variability and spread of the groundnut (Arachis hypogaea) (English translation by Smartt J) In: Ucko RJ, Dimbledy CW (eds) The domestication and exploitation of plants and animals. Duckworth, London, p 427–441

Krapovickas A, Gregory WC (1994) Taxonomia del genero *Arachis* (*Leguminosae*) Bonplandia VIII:1–187. (In Spanish). (English translation by Williams DE, Simpson CE 2007). Taxonomy of the genus Arachis (Leguminosae). Bonplandia 16 (suppl): 1–205

Latham M, Jonsson U, Sterken E, Kent G (2011) RUTF stuff: Can the children be saved with fortified peanut paste? World Nutr 2:62–85

Leal-Bertioli SCM, Jose ACVF, Alves-Freitas DMT, Mortezsohn MC, Guimaraes PM, Nielen S, Vidigal BS, Pereira RW, Pike J, Favero AP, Parniske M, Varshney RK, Bertioli DJ (2009) Identification of candidate genome regions controlling disease resistance in *Arachis*. BMC Plant Biol 9:112

Li Z, Jarret R, Demski J (1997) Engineered resistance to tomato spotted wilt virus in transgenic peanut expressing the viral nucleocapsid gene. Transgenic Res 6:297–305

Liao B (2014) Peanut breeding. In: Mallikarjuna N, Varshney RK (eds) Genetics, genomics and breeding of peanuts. CRC Press, London, New York, pp 61–78

Livingstone DM, Hampton JL, Phipps PM, Grabau EA (2005) Enhancing resistance to *Sclerotinia minor* in peanut by expressing a barley oxalate oxidase gene. Plant Physiol 137:1354–1362

Luo M, Liang XQ, Dang P, Holbrook CC, Bausher MG, Lee RD, Guo BZ (2005) Microarray-based screening of differentially expressed genes in peanut in response to *Aspergillus parasiticus* infection and drought stress. Plant Sci 169:695–703

Mehta R, Radhakrishnan T, Kumar A, Yadav R, Dobaria JR, Thirumalaisamy PP, Jain RK, Chigurupati P (2013) Coat protein-mediated transgenic resistance of peanut (*Arachis hypogaea* L.) to peanut stem necrosis disease through *Agrobacterium*-mediated genetic transformation. Indian J Virol 24:205–213

Moncada S, Higgs A (1993) The L-arginine-nitric oxide pathway. N Engl J Med 329:2002–2012

Murthy TGK, Reddy PS (1993) Genetics of groundnut. In: Murthy TGK, Reddy PS (eds) Cytogenetics and genetics of groundnuts. Antercept, Andover, UK, pp 144–268

Nagaraj G (1988) Chemistry and utilization. In: Reddy PS (ed) Groundnut. Indian council of Agricultural Research, New Delhi, pp 555–565

Nagy ED, Guo Y, Tang S, Bowers JE, Okashah RA, Taylor CA, Zhang D, Khanal S, Heesacker AF, Khalilian N, Farmer AD, Carrasquilla-Garcia N, Penmetsa RV, Cook D, Stalker HT, Nielsen N, Ozias-Akins P, Knapp SJ (2012) A high-density genetic map of *Arachis duranensis*, a diploid ancestor of cultivated peanut. BMC Genom 13:469

Nigam SN (2015) Groundnut at a glance. Feed the future innovation lab for collaborative research on peanut productivity and Mycotoxin control, USAID, p 99

Nigam SN, Upadhyaya HD, Chandra S, Nageswar Rao RC, Wright GC, Reddy AGS (2001) Gene effects for specific leaf area and harvest index in three crosses of groundnut (*Arachis hypogaea* L.). Ann Appl Biol 139:301–306

Olorunju PE, Kuhn CW, Demski JW, Misari SM, Ansa OA (1992) Inheritance of resistance in peanut to mixed infections of groundnut rosette virus (GRV) and groundnut rosette assistor virus and a single infection of GRV. Plant Dis 76:95–100

Ozias-Akins P, Schnall JA, Anderson WF, Singsit C, Clemente TE, Adang MJ, Weissinger AK (1993) Regeneration of transgenic peanut plants from stably transformed embryogenic callus. Plant Sci 93:185–194

Pandey MK, Guo B, Holbrook CC, Janila P, Zhang X, Bertioli DJ, Isobe S, Liang X, Varshney RK (2014) Molecular markers, genetic maps and QTLs for molecular breeding in Peanut. In: Mallikarjuna N, Varshney RK (eds) Genetics, genomics and breeding of peanuts. CRC Press, London, New York, pp 79–113

Pensuk K, Jogloy S, Wongkaew S, Patanothai A (2004) Generation means analysis of resistance to peanut bud necrosis caused by peanut bud necrosis tospovirus in peanut. Plant Breed 123:90–92

Qin H, Gu Q, Zhang J, Sun L, Kuppu S, Zhang Y, Burow M, Payton P, Blumwald E, Zhang H (2011) Regulated expression of an isopentenyltransferase gene (IPT) in peanut significantly improves drought tolerance and increases yield under field conditions. Plant Cell Physiol 52:1904–1914

Qin H, Gu Q, Kuppu S, Sun L, Zhu X, Mishra N, Hu R, Shen G, Zhang J, Zhang Y, Zhu L, Zhang X, Burow M, Payton P, Zhang H (2013) Expression of the Arabidopsis vacuolar H + -pyrophosphatase gene AVP1 in peanut to improve drought and salt tolerance. Plant Biotechnol Rep. 7:345–355

Qin HD, Feng SP, Chen C, Guo YF, Knapp S, Culbreath A, He GH, Wang ML, Zhang XY, Holbrook CC, Ozias-Akins P, Guo BZ (2012) An integrated genetic linkage map of cultivated peanut (*Arachis hypogaea* L.) constructed from two RIL populations. Theor Appl Genet 124:653–664

Rabjohn P, Helm EM, Stanley JS, West CM, Sampson HA, Burks AW, Bannon GA (1999) Molecular cloning and epitope analysis of the peanut allergen Ara h 3. J Clin Invest 103:535–542

Rao MJV, Nigam SN, Mehan VK, McDonald D (1989) *Aspergillus flavus* resistance breeding in groundnut: progress made at ICRISAT Center. In: McDonald D, Mehan VK (eds) Aflatoxin contamination of groundnut. Proc Int Workshop, 6–9 Oct 1987, ICRISAT Center. International crops research institute for the semi-arid tropics, Patancheru, AP, India, p 345–355

Rathnakumar AL, Hariprasanna K, Lalwani HB (2010) Genetic improvement in Spanish type groundnut (*Arachis hypogaea* L.) varieties in India over the years. Ind J Oilseeds Res 27:1–7

Rathnakumar AL, Singh R, Parmar DL, Mishra JB (2013) Groundnut: a crop profile and compendium of notified varieties of India, Directorate of Groundnut Research, Junagadh, Gujarat, p 118

Rathnakumar AL, Nigam SN, Muralidharan V, Mishra JB (2015) Groundnut Scenario. In: Rathnakumar AL, Nigam SN, Muralidharan V, Mishra JB (eds) Groundnut at a cross road in India. Mumbai, India, pp 3–4

Reddy PS (1988) Genetics, breeding and varieties. In: Reddy PS (ed) Groundnut. Indian Council of Agricultural Research, New Delhi, pp 200–317

Reddy PS, Murthy TGK (1996) Current genetic research on groundnut in India. Genetica 97:263–277

Sampson HA (2004) Update on food allergy. J Allergy Clin Immunol 113(5):805–819

Sanders TH, McMichael RW, Hendrix KW (2000) Occurrence of resveratrol in edible peanuts. J Agric Food Chem 48:1243–1246

Sanders TH, Vercellotti JR, Crippen KL, Hinsch RT, Rasmussen GK, Edwards JH (1992) Quality factors in exported peanuts from Argentina, China and the United States. J Amer Oil Chem Soc 69(10):1032–1035

Seijo GJ, Lavia GI, Fernandez A, Krapovickas A, Ducasse AD, Moscone EA (2004) Physical mapping of the 5 S and 18S-25 S rRNA genes by FISH as evidence that *Arachis duranensis* and *A. ipaënsis* are the wild diploid progenitors of A. hypogaea (Leguminosae). Am J Bot 91:1294–1303

Seijo GJ, Lavia GI, Fernandez A, Krapovickas A, Ducasse AD, Bertioli DJ, Moscone EA (2007) Genomic relationships between the cultivated peanut (*Arachis hypogaea*, Leguminosae) and its close relatives revealed by double GISH. Am J Bot 94:1963–1971

Sharma SKK, Anjaiah VV (2000) An efficient method for the production of transgenic plants of peanut (*Arachis hypogaea* L.) through *Agrobacterium* tumefaciens-mediated genetic transformation. Plant Sci 159:7–19

Singh AK, Mehan VK, Nigam SN (1997) Sources of resistance to groundnut fungal and bacterial diseases: an update and appraisal. Information Bulletin no. 50. International Crops Research Institute for the Semi-Arid Tropics, Patancheru, AP, India, p 48

Singh AK, Subrahmanyam P, Moss JP (1984) The dominant nature of resistance to *Puccinia arachidis* in certain wild *Arachis* species. Oleagineux 39:535–538

Tiwari SP, Ghewande MP, Misra DP (1984) Inheritance of resistance to rust and late leaf spot in groundnut (*Arachis hypogaea* L.). J Cytol Genet 19:97–101

Tiwari S, Mishra DK, Singh A, Singh PK, Tuli R (2008) Expression of a synthetic cry1EC gene for resistance against *Spodoptera litura* in transgenic peanut (*Arachis hypogaea* L.). Plant Cell Rep 27:1017–1025

Upadhyaya HD, Sharma S, Dwivedi SL (2011) Arachis. In: Kole C (ed) Wild crop relatives: genomic and breeding resources: legume crops and forages. Springer publishing house, London, New York, pp 1–20

Upadhyaya HD, Nigam SN, Mehan VK, Lenne JM (1997) Aflatoxin contamination of groundnut—prospects of a genetic solution through conventional breeding, p. 81–85. In: Aflatoxin Contamination Problems in Groundnut in Asia: Proceedings of the First Working Group Meeting, 27–29 May 1996, Ministry of Agriculture and Rural Development, Hanoi, Vietnam (V.K. Mehan and C.LL. Gowda eds.). Patancheru 502 324, Andhra Pradesh, India: International Crops Research Institute for the Semi-Arid Tropics

Utomo SD, Anderson WF, Wynne JC, Beute MK, Hagler WM Jr, Payne GA (1990) Estimates of heritability and correlation among three mechanisms of resistance to *Aspergillus parasiticus* in peanut. Proc Amer Peanut Res and Educ Soc 22:26

Valls JFM, Simpson CE (2005) New species of *Arachis* (Leguminosae) from Brazil, Paraguay and Bolivia. Bonplandia 14:35–63

Varshney RK, May GD (2012) Next-generation sequencing technologies: opportunities and obligations in plant genomics. Brief Funct Genom 11:1–2

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 (2009) The first SSR-based genetic linkage map for

cultivated groundnut (*Arachis hypogaea* L.). Theor Appl Genet 118:729–739

Woyengo TA, Ramprasath VR, Jones PJ (2009) Anticancer effects of phytosterols. Eur J Clin Nutr 63 (7):813–820

Yang H, Singsit C, Wang A, Gonsalves D, Ozias-Akins P (1998) Transgenic peanut plants containing a nucleocapsid protein gene of tomato spotted wilt virus show divergent levels of gene expression. Plant Cell Rep 17:693–699

Yen GC, Duh PD (1994) Scavenging effects of methanolic extracts of peanut hulls on free-radical and active-oxygen species. J Agri Food Chem 42 (3):629–632

Yocum MW, Khan DA (1994) Assessment of patients who have experienced anaphylaxis: a 3-year survey. Mayo Clin Proc 69:16–23

Yu SL (2011) Peanut genetics and breeding in China. Shanghai Scientific and Technology Press, Shanghai, China

Zhou GY, Liang XQ (2002) Analysis of major-minor genes related to resistance to infection by *Aspergillus flavus* in peanut. J Peanut Sci 31:11–14

Zhou GY, Liang XQ, Li YC, Li XC, Li SL (1999) Evaluation and application of introduced peanut cultivars for resistance to Aspergillus flavus invasion. J Peanut Sci 32:14–17

Zhou XZ, Xia Y, Ren X, Chen Y, Huang L, Huang S, Liao B, Lei Y, Yan L, Jiang H (2014) Construction of an SNP based genetic linkage map in cultivated peanut based on large scale marker development using next-generation double-digest restriction site associated DNA sequencing (ddRADseq). BMC Genom 15:351

# Peanut: Origin and Botanical Descriptions

3

Shyam P. Tallury

**Abstract**

Since the first description of the cultivated peanut, *Arachis hypogaea* L. by Linneaus in 1753, to the recent monograph on the taxonomy of genus *Arachis* (Krapovickas and Gregory in Bonplandia 8(1–4):1–186, 1994; Krapovickas and Gregory in Bonplandia 16(Supl.):1–205, 2007), our knowledge of the genetic structure of the genus including its origin, variability, and geographical distribution of various species has significantly improved. Large germplasm collections have been accumulated in several countries to preserve the genetic diversity and characterize the germplasm resources for efficient utilization in peanut improvement programs. Plant growth and development including the origin and botanical classification of the cultivated species have been revisited here to summarize our current knowledge of the taxon. In spite of these advances, taxonomic and phylogenetic ambiguities still exist. It is likely that the recent advances in peanut genome sequencing and the availability of new and additional molecular markers and other genomic tools might help clarify the genetic structure of genus *Arachis* and of the cultivated species, *A. hypogaea*.

## 3.1 Introduction

Peanut (*Arachis hypogaea* L.), is an important grain legume crop and is primarily valued as a source of protein as well as fat to human nutrition. The seeds contain about 20–25% protein and 45–55% oil, in addition to useful vitamins and minerals, and offer an easily affordable source of protein for many, particularly in the developing countries. Of the 81 described species, *A. hypogaea*, is the only domesticated species and is cultivated for its seeds for human consumption, although a few other species have been reported to have uses for nutrition, forage and ornamental value (Krapovickas and Gregory 1994, 2007; Gimenes et al. 2000; Simpson et al. 2001; Galgaro et al. 1997; Stalker and Simpson

S.P. Tallury (✉)
Plant Genetic Resources Conservation Unit, USDA-ARS, 1109 Experiment Street, Griffin GA 30223-1797, USA
e-mail: Shyam.Tallury@ars.usda.gov


R.K. Varshney et al. (eds.), *The Peanut Genome*, Compendium of Plant Genomes, DOI 10.1007/978-3-319-63935-2_3

1995). *Arachis hypogaea* is a herbaceous annual with plants of about a 45–60 cm tall and 30 cm wide with a deep taproot. The roots typically contain the nitrogen fixing bacterial nodules of *Rhizobium*, which coexist in a symbiotic relationship by providing the necessary nitrogen for plant growth while deriving sustenance from it. The plants are self-fertilizing and have a unique mode of reproduction where flowers are produced on the plant and following fertilization, the gynoecium enters the soil through formation of a peg. Then, pods containing the seeds are produced underground. All species in the genus produce underground pods and in a sense are "pegged" to the ground as mentioned in Krapovickas and Gregory (1994, 2007). Because of this unique reproductive feature, peanut is also known as groundnut in many parts of the world.

### 3.1.1 Origin and Distribution of Genus *Arachis*

*Arachis* is a native South American genus with natural populations found growing in Argentina, Bolivia, Brazil, Paraguay, and Uruguay (Valls et al. 1985). The genus likely originated in the highlands of southwestern Mato Grosso do Sul state in Brazil (Hammons 1973; Gregory et al. 1980; Simpson et al. 2001) where the most ancient, trifoliate species, *A. guaranitica* Chodat. and Hassl., and *A. tuberosa* Bong. Ex Benth. were collected. *Arachis guaranitica* is the most genetically isolated species and looks more like a grass plant. These two species are still found growing in this region (Simpson et al. 2001). Subsequently, with water movement, the species spread to drier lowlands in all directions and evolved into various river valleys and drainage systems (Gregory and Gregory 1979; Stalker and Simpson 1995; Simpson et al. 2001) with *Arachis* species growing in sandy to heavy clay/loamy soils and on schist rocks with no soil (Simpson et al. 2001). One of the species, *A. burkartii* Handro was collected in southern Brazil in black gummy clay mixed with small stones with a soil pH of 3.2 (Stalker and Simpson 1995), indicating the wide adaptation of *Arachis* species to extremely diverse geographical environments. The geocarpic reproductive development probably protected the pods/seeds from the predators and helped in sustained survivability and distribution of the genus in South America. However, it is also possible that the geocarpic pod limited the rapid spread of the genus as estimated by Simpson et al. (2001) that the species moved only one meter/year across the continent.

Currently, the genus contains about 81 described species and several new species are likely to be described in the near future (Stalker et al. 2016; Simpson, personal communication). Krapovickas and Gregory (1994, 2007) delineated the species diversity into nine different sections based on geographical distributions, plant, pod and chromosome morphologies and cross-compatibility relationships. The cultivated species, *A. hypogaea*, was assigned to section *Arachis*, which also contains a number of wild species. They concluded that *A. hypogaea* hybridizes readily with the species in section *Arachis* whereas the species in the remaining eight sections are incompatible with it. Although the genus *Arachis* originated in the highlands of Brazil, the center of origin of the cultivated species, *A. hypogaea*, is believed to be southern Bolivia to northwestern Argentina. This observation was based on the presence of the parental diploid wild species of *A. hypogaea* in this region, the wide range of variation observed in pod and seed morphologies and that the germplasm collected in this area exhibited primitive characters associated with wild species, thus supporting the likely origin of *A. hypogaea* in this region (Hammons 1982; Stalker and Simpson 1995; Ferguson et al. 2004). Additionally, Simpson et al. (2001) suggested possible alternate regions for the origin of *A. hypogaea* on the west coast of Peru and/or the eastern slopes of Cordillera in the Andes, based on archeological evidence and prevalence of favorable environmental conditions for survival of plants for long periods of time.

*Arachis hypogaea* is an allotetraploid ($2n = 4x = 40$) with a genomic composition of

AABB. It is considered to have originated from natural hybridization of two closely related diploid wild species followed by either chromosome duplication or fusion of unreduced gametes, thereby, resulting in an allotetraploid with two sets of chromosomes of each of the constituent parental genomes (Seijo et al. 2004, 2007). The general agreement among peanut researchers is that *A. duranensis* Krapov. & W.C. Gregory and *A. ipaënsis* Krapov. & W.C. Gregory are the A and B genome donor species, respectively (Kochert et al. 1991, 1996; Seijo et al. 2004, 2007). Recently, Bertioli et al. (2016) reported the genome sequences of these two species and demonstrated that they are very similar to the A and B subgenomes of *A. hypogaea*.

A list of primitive and advanced traits was compiled by Stalker and Simpson (1995) to demonstrate the evolution of domesticated peanut. Krapovickas (1968) suggested that *A. hypogaea* subsp. *hypogaea* var. *hypogaea* (see Sect. 1.1.2 below for the subspecies descriptions) was the most ancient cultivar type. His observation was based on the available records that it was the most predominant type found in the chaco region between southern Bolivia and northwestern Argentina, the likely area where *A. hypogaea* is believed to have originated. Additionally, the types found in this area exhibited many primitive traits such as the runner growth habit, a branching pattern similar to the wild *Arachis* species, small, two-seeded pods with marked constriction and slight reticulation, and seed dormancy. The above observations led Krapovickas and Gregory (1994, 2007) to conclude that south east Bolivia is the center of origin as well as diversity for subsp. *hypogaea*, whereas subsp. *fastigiata* differentiated in north western Bolivia and possibly in Peru, along with vars. *fastigiata*, *peruviana* and *aequatoriana*. However, genetic diversity analysis among botanical varieties using simple sequence repeat (SSR) markers by Ferguson et al. (2004) revealed similarities of three botanical varieties of subsp. *fastigiata*, namely *fastigiata*, *vulgaris* and *aequatoriana* but did not support the inclusion of var. *peruviana* in subsp *fastigiata*. Further, they also found that the botanical varieties, *hypogaea,* and *hirsuta* were not closely related and suggested that they should not be grouped under subsp. *hypogaea*. Contrarily, He and Prakash (2001) demonstrated with AFLP markers that vars. *aequatoriana* and *peruviana* were closer to subsp. *hypogaea* than to subspecies *fastigiata*. Thus, there still exists, considerable confusion about the taxonomic classification of the cultivated species. Among the market types, Gregory et al. (1980) and Hammons (1982) suggested that the Bolivian and Amazonian geographic regions are the possible sites for the origin of the large-seeded Virginia types. Further, Hammons (1982) indicated that the Guarani area of northeastern Argentina, Paraguay and southern Brazil is the center of variation for the Spanish (var. *vulgaris*) market type whereas, the Valencia type (var. *fastigiata*) probably spread from Paraguay and central Brazil (Hammons 1982; Krapovickas 1968).

Further, Krapovickas (1968) and Gregory and Gregory (1976) recognized six other regions in South America as the secondary centers of diversity for the cultivated species based on morphological variability of the landraces. Additionally, Africa, China and India are considered as tertiary centers of diversity for *A. hypogaea* because of the large number of landraces and other local germplasm displaying different pod and seed traits (Gibbons et al. 1972). To characterize and describe the vast amount of morphological variation present in *A. hypogaea*, peanut descriptor lists were compiled (IBPGR and ICRISAT 1992; Pittman 1995). These descriptors included a standardized set of plant, pod, and seed traits to help classify the cultivated germplasm into related groups. The USDA National Plant Germplasm System peanut collection maintained at the Plant Genetic Resources Conservation Unit (PGRCU) in Griffin, GA, USA, routinely uses the U S peanut descriptors developed by Pittman (1995) to characterize the collection. Digital images of the various plant, pod and seed features are also compiled. This information is made available to

researchers around the world on the Germplasm Resources Information Network Global at www.ars-grin.gov site.

It is paradoxical that in spite of the extensive morphological variation among the subspecies and botanical varieties of *A. hypogaea*, little molecular (DNA) polymorphism was observed in the cultivated species (Kochert et al. 1991; Halward et al. 1991, 1992; Moretzsohn et al. 2004, 2013; Pandey et al. 2012). A likely hypothesis for the lack of molecular polymorphisms in the cultivated species was that a single hybridization event accompanied by polyploidization coupled with the autogamous reproduction led to the genetic isolation of the raw tetraploid from the surrounding species diversity with no apparent gene flow between them (Kochert et al. 1996; Seijo et al. 2007). It is likely that following domestication, the early humans selected desirable types from the original population possibly for compact habit, and increased pod and seed sizes producing the different subspecies and varieties of the cultivated taxon, as we have today. Consequently, the vast amount of morphological variability observed in the cultivated taxon is likely to have resulted from natural and/or artificial selection rather than from the introgression of genes from different species (Seijo et al. 2007). However, Varshney et al. (2009) using simple sequence repeat (SSR) markers with a diverse set of 189 *A. hypogaea* accessions observed significant polymorphisms and grouped the accessions into four different clusters. It is also encouraging that the recent advances in peanut genome sequencing and new genomic tools might help clarify the origin, evolution, variability and distribution of the genus and that of the cultivated species, *A. hypogaea*. An initial application of these technologies, particularly, of molecular markers for quantitative trait analysis was demonstrated by Pandey et al. (2012, 2013) for use in marker assisted breeding in cultivated peanut.

Contrarily, the wild species have exhibited extensive molecular variation among and within the different sectional groups (Halward et al. 1991, 1992; Tallury et al. 2005; Upadhyaya et al. 2008a, b; Friend et al. 2010; Moretzsohn et al. 2013). Molecular profiling of a composite collection consisting of 1000 diverse peanut accessions which included both cultivated and wild species demonstrated rich allelic diversity within the wild species with more than 100 unique alleles (Upadhyaya et al. 2008a, b) whereas the number of unique alleles in the two *A. hypogaea* subspecies, *hypogaea* and *fastigiata* were only 11 and 50, respectively. Further, the highest number of unique alleles were found in *A. hypogaea* accessions from the Americas with few unique alleles among the accessions from Asia and Africa. This study also demonstrated that the two subspecies, *hypogaea* and *fastigiata* accessions shared 70 alleles among them. Although the wild species shared only 15 alleles with subspecies *hypogaea* and 32 alleles with subspecies *fastigiata*, the wild species accessions grouped with subspecies *hypogaea* accessions (Upadhyaya et al. 2008a, b).

#### 3.1.1.1 Geographical Spread *of A. hypogaea*

Following the Spanish and Portuguese explorations to South America, the cultivated peanut spread from the centers of origin and diversity in South America to Europe and then to Africa and Asia via trade voyages. There is no substantiated evidence for the occurrence of cultivated peanut in North America during this time. It was suggested that peanut was introduced into U.S. on slave trade ships from Africa via the coast of northeastern Brazil, where peanut was gathered as food source to complete the journey, strongly suggesting that the first peanut introductions into the U.S. were from Brazil rather than from Africa (Stalker and Simpson 1995).

#### 3.1.1.2 Botanical Classification of *A. hypogaea*

Krapovickas and Gregory (1994, 2007) indicated that genus *Arachis* is defined by its morphological features of the underground structures, including the pods, rhizomatous stems, root systems, and hypocotyls. They showed that these defining characters grouped the *Arachis* collections into different geographic areas and ecological features. This, along with crossabilities of species, allowed them to group the collections

into nine different sections (Gregory and Gregory 1979; Krapovickas and Gregory 1994, 2007). *Arachis hypogaea* belongs to section *Arachis*, which also contains 30 other wild species.

Further, *Arachis hypogaea* was divided into two subspecies, subsp. *hypogaea* and subsp. *fastigiata* by Krapovickas and Rigoni (1960) based on the absence versus presence of flowers on the main stem. They also proposed two botanical varieties of subsp. *fastigiata*, vars. *fastigiata* and *vulgaris* based on the pod traits. Later, Krapovickas (1968) proposed that subsp. *hypogaea* should also be divided into vars. *hypogaea* and *hirsuta*. With additional collections of *A. hypogaea*, Krapovickas and Gregory (1994, 2007) not only confirmed the two subspecies of *A. hypogaea* (subsp. *hypogaea* and subsp. *fastigiata*) but also expanded botanical varietal groups to six (vars. *hypogaea, hirsuta, fastigiata, peruviana, aequatoriana and vulgaris*) based on plant growth habit, leaf color and branching patterns as described below, which also includes the four major market types grown in the U.S.

A.
***A. hypogaea* subsp. *hypogaea* L.**
No flowers on main stem
Alternating pairs of floral and vegetative axes on branches
Branches short and less hairy
Dark green leaves
Prostrate or spreading growth habit
Late maturing
Seed dormancy present

**var. *hypogaea***:

Leaflets with glabrous dorsal surface; with a few hairs on the midrib
Prostrate/spreading (runner) or bunch type growth habit

**Market type: Virginia**

Less hairy short main stem and leaves
Large pods, two seeded
Slight constriction and reticulation

**Market type: Runner**

Less hairy main stem and leaves
Small pods, two seeded
Slight constriction and reticulation

**var. *hirsuta***:

Leaflets with 1–2 mm long hairs dispersed on entire dorsal surface
Long main stem and very hairy,

**Market type: Peruvian runner**

More hairy leaves
Late maturing
Long pods, 2-3 seeded
Deep constriction and prominent reticulation

B.
***A. hypogaea* subsp. *fastigiata***
Flowers on main stem
Sequential floral and vegetative branches
Branches less hairy
Light green leaves
Bunch or erect growth habit
Early maturing
Seed dormancy absent

**var. *fastigiata***

Leaflets with glabrous dorsal surface or hairs only on the midrib
Few branches, short and slender
Pods with smooth or slight reticulation

**Market type: Valencia**

Sparsely branched; curved branches
Erect growth habit
Usually 2–4 seeded, long pods
Red seed coat

**var. *aequatoriana***

Erect plants with large leaves

Leaflets 1–2 mm long. hairy dorsal surface, dispersed on entire surface
Main stem with short inflorescences
Long reproductive lateral branches
Prominent longitudinal ribs on pods with deep pod reticulation
Long pods with 3–4 seeded
Seed coat is commonly violet

**var. *peruviana***

Thick, large leaves; leaflets glabrous on both sides
Hairy on the margins and dorsal surface on midrib of leaflets
Long, robust reproductive branches
Flowers on both main stem and lateral branches
3–4-seeded pods
Seed coat colors vary from black, violet, cream to variegated
Prominent longitudinal ribs on pods with deep pod reticulation

**var. *vulgaris***

Erect growth habit with many upright branches
Medium sized leaves with glabrous surface, long hairs on margins
Mostly 2-seeded, small pods bunched at the base of the plant
Slight pod constriction and reticulation

**Market type: Spanish**

More branched; upright branches
Light green leaves

## 3.2 Description of Seed to Adult Plant

### 3.2.1 Germination and Plant Morphology

*Arachis hypogaea* seed is covered by a thin seed coat or testa (commonly called as "skin"). It varies in color from white to tan to black and different shades of red or pink and provides protection to the seed from soil microorganisms. The seeds also vary in size from the large Virginia market types to the small, rounded Spanish types. The seed is composed of two cotyledons, which are the first true leaves. They contain stored food reserves for the young seedling during germination and to establish a plant. The peanut seed contains the dormant shoot (plumule/leaf primordia) and the root initials (radicle). When a seed is planted under optimum soil moisture and temperature conditions, the process of germination is initiated in about three to five days. First, the radicle starts to grow forming the upper hypocotyl and the lower primary root. This is accompanied by the rapid elongation of the hypocotyl which pushes the cotyledons above ground within a week. The cotyledons split open to expose the shoot primordia which extends to form the epicotyl. The epicotyl forms the main stem whereas the lower hypocotyl elongates to form the tap root (Gregory et al. 1973). From the taproot, lateral roots emerge within 3–5 days after germination and are extensively developed by about 7–10 days. Occasionally, on mature plants, adventitious roots are formed where branches are in contact with soil.

Peanut plant is an erect or prostrate type where plants are usually about 30–45 cm tall and the lateral branches spread to about 30 cm wide. However, on many wild species, the lateral branches are a few meters long with a very short main stem. Plants have compound leaves with four leaflets (tetrafoliate) which are located alternately on the main stem and lateral branches. However, three species from the section Triserectoides, namely, *A. guaranitica A. tuberosa* and *A. sesquijuga* have leaves with three leaflets (trifoliate). The leaves are connected to the stems by an adnate stipule and leaflets vary in size and shape, usually oblong and lanceolate in some wild species. The *A. hypogaea* subsp. *hypogaea* has dark green leaves compared to the lighter green leaves in *A. hypogaea* subsp. *fastigiata*. The stems are angular, mostly green with the exception of Valencia and *aequatoriana* types which are reddish purple. As summarized in the

previous section, in some botanical varieties hairy stems are commonly seen. Stem pigmentation, hairiness on stems and leaves has been shown to deter leaf feeding insect pests (Campbell et al. 1976; Stalker and Campbell 1983; Stalker et al. 1984; Sharma et al. 2003).

### 3.2.2 Flower Morphology, Fertilization and Pod/Seed Development

Generally, peanut plants produce flowers about 30 days after germination. It is an indeterminate plant and as a result, flowers are produced throughout the growing season leading up to harvest. Flowers are usually present in leaf axils on the branches and also on the main stem in subspecies *fastigiata* types. Commonly, three flowers are present in each inflorescence, which is a raceme. At any given time, usually only one flower opens and the interval between the openings of flowers in the same inflorescence vary up to several days. Since the flowers contain fertile, male and female reproductive parts, natural self-fertilization leads to the development of pods.

The flower has five brightly colored petals (corolla) consisting of a large standard (Banner), two wing petals and two fused keel petals. The calyx is green with five lobes of which, one lobe is opposite the keel whereas the other four are fused and cover the back side of the standard. The standard is usually yellow or orange with red veins on the inner face. The wings are usually yellow surrounding the keel. The keel is almost colorless and encloses the stamens and style. The androecium is monadelphous with filaments fused for two-thirds of their length and contains eight functional stamens and two, small sterile ones. The stamens produce pollen (male gamets) for fertilization of the egg cell. The flower is attached to the stem (at the leaf axil) by a long tube like structure called a hypanthium or "calyx tube". The flowers are subtended by a bract and are sessile although they appear as pedicillate because of tubular hypanthium. The style is enclosed within the hypanthium and is connected to the ovary (female part) located at the base of the hypanthium in the leaf axil. The tip of the style, called stigma is at the same level or slightly above the anthers so pollen grains can reach it. Differences in stigma morphology were observed between *A. hypogaea* and the wild species. In *A. hypogaea*, the stigma is of dry papillate type (Lakshmi and Shivanna 1986) with no surrounding hairs and probably accommodates about 15 pollen grains (Moss and Rao 1995). On the other hand, the annual *Arachis* species have large stigmatic surface whereas the perennial species have smaller, cuticularized stigmas with unicellular hairs accommodating a maximum of only three pollen grains (Lu et al. 1990). However, in the wild species, *A. lignosa*, the truncated shape of the stigma and its elevated position relative to the anthers restricts natural self-pollination (Banks 1990). In this case, manual tripping of flowers is needed for pollen to reach stigma for fertilization and later pod development. Outcrossing is possible with bees or other insects, however, it is limited to less than 10% under natural field conditions (Hammons 1973; Knauft et al. 1992).

The process of fertilization begins with anthesis, which occurs within a few hours after sunrise with the opening of the flower. The mature pollen grain is two-celled with two generative nuclei (Xi 1991). The ovary usually has two ovules, and up to three or more in Valencia types. Each ovule contains a mature embryo sac with a well-differentiated egg cell at the micropylar end and a polar nucleus surrounded by starch grains. When pollen germinates on a receptive stigma, the pollen tube containing the mature pollen grain (male gamete) with the two generative nuclei travels through the style and eventually enters the embryo sac through the micropyle. One of the two generative nuclei fuses with the egg cell (syngamy) to form the embryo and the other with the polar nucleus (double fertilization) to form the endosperm. Following syngamy, the starch grains breakdown to provide initial nutrition for the proembryo to grow which eventually develops into a mature seed. The entire process of fertilization usually

takes about 18–24 h from anthesis to syngamy (Pattee et al. 1991). Each ovule develops into a peanut seed and the ovary becomes the pod.

Following syngamy, pod development is initiated as a pointed, stalk like structure which is called the "peg" (Smith 1950). During the early embryo growth (24–72 h after fertilization), an intercalary meristem at the base of the ovary actively divides forming the peg with the fertilized ovules at its tip. Pegs are usually observed between 4 and 7 days after self-pollination and are positively geotropic (Zamiski and Ziv 1976) and require darkness for pod formation (Ziv 1981). As the peg is extending to enter the soil (aerial phase), the embryo remains in a quiescent stage, usually, as an 8-celled proembryo (Pattee and Mohapatra 1987). It is only after soil penetration that the elongation of the peg is arrested, to initiate pod formation. The first sign of pod development involves the swelling of the peg tip along with the horizontal turning of the peg. The peg becomes diageotropic after soil penetration such that the ovules are always located on the upper wall of the pod, with the pod tip pointing away from the plant (Moss and Rao 1995). Pod enlargement occurs from base towards the tip with simultaneous faster development of the basal ovule (Smith 1950). A mature peanut pod is developed in about 60–80 days after pollination. However, due to the indeterminate nature of peanut plants, flowering continues throughout the growing season until harvest. Consequently, pods at different maturities are seen on plants even at harvesting time. Detailed descriptions of peanut embryology including the growth and development of pegs, pods, and seeds are well documented in literature (Smith 1950; Gregory et al. 1973; Periasamy and Smapoornam 1984; Pattee and Mohapatra 1987; Xi 1991; Moss and Rao 1995).

## 3.3 Conclusion

Plants of genus *Arachis* are characterized by their unique underground structures, including the root systems, rhizomatous stems, pods, and hypocotyls. These features led to the adaptation and grouping of *Arachis* germplasm into different geographical regions and evolution of botanical varieties. *Arachis hypoagea* is a native, new world taxon and exhibits large morphological variation as described above with a wide range of adaptation to many different ecological conditions. An understanding of preservation and characterization of this genetic diversity is crucial to future genetic improvement of *A. hypogaea*.

## References

Banks DJ (1990) Hand-tripped flowers promote seed production in *Arachis lignosa*, a wild peanut. Peanut Sci 17:23–24

Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EKS, Liu X, Gao D, Clevenger J, Dash S, Ren L, Moretzsohn MC, Shirasawa K, Wei W, Vidigal B, Abernathy B, Chu Y, Niederhuth CE, Umale P, Araújo ACG, Kozik A, Kim KD, Burow MD, Varshney RK, Wang X, Zhang X, Barkley N, Guimarães PM, Isobe S, Guo B, Liao B, Stalker HT, Schmitz RJ, Scheffler BE, Leal-Bertioli SCM, Xun X, Jackson SA, Michelmore R, Ozias-Akins P (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet 48:438–446

Campbell WV, Emery DA, Wynne JC (1976) Resistance of peanuts to the potato leafhopper. Peanut Sci 3:40–43

Ferguson ME, Bramel PJ, Chandra S (2004) Gene diversity among botanical varieties in peanut (*Arachis hypogaea* L.). Crop Sci 44:1847–1854

Friend SA, Quandt D, Tallury SP, Stalker HT, Hilu KW (2010) Species, genomes and section relationships in genus *Arachis* (Fabaceae): a molecular phylogeny. Plant Syst Evol 290:185–199

Galgaro L, Valls JFM, Lopes CR (1997) Study of the genetic variability and similarity among and within *Arachis villosulicarpa*, *A. pietrarellii* and *A. hypogaea* through isoenzyme analysis. Genet Resour Crop Evol 44:9–15

Gibbons RW, Bunting AH, Smartt J (1972) The classification of varieties of groundnut (*Arachis hypogaea* L.). Euphytica 21:78–85

Gimenes MA, Lopes CR, Galgaro ML, Valls JFM, Kochert G (2000) Genetic variation and phylogenetic relationships based on RAPD analysis in section Caulorrhizae, genus *Arachis* (Leguminosae). Euphytica 116:187–195

Gregory WC, Gregory MP (1976) Groundnut. In: Simmonds NW (ed) Evolution of crop plants. Longman Group Ltd., London, pp 151–154

Gregory MP, Gregory WC (1979) Exotic germplasm of *Arachis* L. interspecific hybrids. J Hered 70:185–193

Gregory WC, Gregory MP, Krapovickas A, Smith BW, Yarbrough JA (1973) Structure and genetic resources of peanuts. Peanuts—Culture and Uses. American Peanut Research and Education Association, Stillwater, OK, pp 47–133

Gregory WC, Krapovickas A, Gregory MP (1980) Structure, variation, evolution and classification in *Arachis*. In: Summerfield RJ, Bunting AH (eds) Advances in Legume Science. Royal Botanic Gardens, Kew, London, pp 469–481

Halward TM, Stalker HT, Larue EA, Kochert G (1991) Genetic variation detectable with molecular markers among unadapted germplasm resources of cultivated peanut and related wild species. Genome 34:1013–1020

Halward T, Stalker T, LaRue E, Kochert G (1992) Use of single-primer DNA amplifications in genetic studies of peanut (*Arachis hypogaea* L.). Plant Mol Biol 18:315–325

Hammons RO (1973) Genetics of *Arachis hypogaea*. Peanuts-culture and uses. American Peanut Res Educ Assn, Stillwater, OK, pp 135–173

Hammons RO (1982) Origin and early history of the peanut. In: Pattee HE, Young CT (eds) Peanut Science and Technology. American Peanut Research and Education Society, Yoakum, TX, pp 1–20

He G, Prakash C (2001) Evaluation of genetic relationships among botanical varieties of cultivated peanut (*Arachis hypogaea* L.) using AFLP markers. Genet Resour Crop Evol 48:347–352

IBPGR and ICRISAT (1992) Descriptors for groundnut (second revision). IBPGR, Rome and ICRISAT, Patancheru, India

Knauft DA, Chiyembekeza AJ, Gorbet DW (1992) Possible reproductive factors contributing to outcrossing in Peanut (*Arachis hypogaea* L.). Peanut Sci 19:29–31

Kochert G, Halward T, Branch WD, Simpson CE (1991) RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. Theor Appl Genet 81:565–570

Kochert G, Stalker HT, Gimenes M, Galgaro L, Lopes CR, Moore K (1996) RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). Am J Bot 83:1282–1291

Krapovickas A (1968) The origin, variability and spread of the groundnut (*Arachis hypogaea*. In: Ucko RJ, Dimbleby CW (eds) The domestication and exploitation of plants and animals. Duckworth, London, pp 427–441

Krapovickas A, Gregory WC (1994) Taxonomia del genero *Arachis* (Leguminosae). Bonplandia 8(1–4):1–186

Krapovickas A, Gregory WC (2007) Taxonomy of the genus *Arachis* (Leguminosae) (trans: Williams DE and Simpson CE). Bonplandia 16(Supl.):1–205

Krapovickas A, Rigoni VA (1960) La nominclatura de las subspecies y variedades de *Arachis hypogaea* L. Revista de Investigaciones Agricoles 14:197–228

Lakshmi KV, Shivanna KR (1986) Structure and cytochemistry of the pistil in *Arachis hypogaea*. Proc Ind Acad Sci (Plant Sci) 95:357–363

Lu J, Mayer A, Pickersgill B (1990) Stigma morphology and pollination in *Arachis* L. (Leguminosae). Ann Bot 66:73–82

Moretzsohn MC, Gouvea EG, Inglis PW, Leal-Bertioli SCM, Valls JFM, Bertioli DJ (2013) A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. Ann Bot 111:113–126

Moretzsohn MC, Hopkins MS, Mitchell SE, Kresovich S, Valls JFM, Ferreira ME (2004) Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. BMC Plant Biol 4:11

Moss JP, Rao VR (1995) The peanut-reproductive development to plant maturity. Advances in peanut science. American Peanut Res Educ Soc, Stillwater, Oklahoma, pp 1–13

Pandey MK, Monyo E, Ozias-Akins P, Liang X, Guimarães P, Nigam SN, Upadhyaya HD, Janila P, Zhang X, Guo B (2012) Advances in *Arachis* genomics for peanut improvement. Biotechnol Adv 30:639–651

Pandey MK, Upadhyaya HD, Rathore A, Vadez V, Sheshasayee MS, Sriswathi M, Govil M, Gowda MVC, Kumar VA, Khera P, Bhat RS, Monyo E, Varshney RK (2013) Association mapping for resistance to abiotic and biotic stresses and agronomically important traits in peanut (*Arachis hypogaea* L.). In: Sixth International Conference of the Peanut Research Community on Advances in *Arachis* through Genomics and Biotechnology. Zhengzhou, China, p 50 (abstract)

Pattee HE, Mohapatra SC (1987) Anatomical changes during ontogeny of the peanut (*Arachis hypogaea* L.) fruit: mature megagametophyte through heart shaped embryo. Bot Gaz 148:156–164

Pattee HE, Stalker HT, Giesbrecht FG (1991) Comparative peg, ovary and ovule ontogeny of selected cultivated and wild-type *Arachis* species. Bot Gaz 152:64–71

Periasamy K, Sampoornam C (1984) The morphology and anatomy of ovule and fruit development in *Arachis hypogaea* L. Ann Bot 53:399–411

Pittman RN (1995) United States peanut descriptors. USDA-ARS-132, US Government printing office, Washington, DC, pp 18

Seijo G, Lavia GI, Fernandez A, Krapovickas A, Ducasse DA, Bertioli DJ, Moscone EA (2007) Genomic relationships between the cultivated peanut (*Arachis hypogaea*, Leguminosae) and its close relatives revealed by double GISH. Am J Bot 94:1963–1971

Seijo G, Lavia GI, Fernandez A, Krapovickas A, Ducasse DA, Moscone EA (2004) Physical mapping of 5S and 18S-25S rRNA genes as evidence that *Arachis duranensis* and *A. ipaensis* are the wild

diploid progenitors of *A. hypogaea* (Leguminosae). Am J Bot 91:1294–1303

Sharma HC, Pampapathy G, Dwivedi SL, Reddy LJ (2003) Mechanisms and diversity of resistance to insect pests in wild relatives of groundnut. J Econ Ent 96:1886–1897

Simpson CE, Krapovickas A, Valls VFM (2001) History of *Arachis* including evidence of *A. hypogaea* L. progenitors. Peanut Sci 28:78–80

Smith BW (1950) *Arachis hypogaea*. Aerial flower and subterranean fruit. Am J Bot 37:802–815

Stalker HT, Campbell WV (1983) Resistance of wild species of peanut to an insect complex. Peanut Sci 10:30–33

Stalker HT, Campbell WV, Wynne JC (1984) Evaluation of cultivated and wild peanut species for resistance to the lesser corn stalk borer, *Elasmopalpus lignosellus* (Lepidoptera: Pyralidae). J Econ Ent 77:53–57

Stalker HT, Simpson CE (1995) Genetic resources in *Arachis*. In: Pattee HE, Stalker HT (eds) Advances in peanut science. Amer Peanut Res Educ Soc, Stillwater, OK, pp 14–53

Stalker HT, Tallury SP, Seijo GR, Leal-Bertioli SC (2016) Biology, speciation and utilization of peanut species. In: Stalker HT, Wilson RF (eds) Peanuts: Genetics. APRES and AOCS Press, Processing and Utilization, pp 27–66

Tallury SP, Hilu KW, Milla SR, Friend SA, Alsaghir M, Stalker HT, Quandt D (2005) Genomic affinities in *Arachis* section *Arachis* (Fabaceae): molecular and cytogenetic evidence. Theor Appl Genet 111:1229–1237

Upadhyaya HD, Bhattacharjee R, Hoisington DA, Chandra S, Varshney RK, Valls JFM, Moretzsohn MC, Leal-Bertioli S, Guimaraes P, Bertioli D (2008a) Molecular characterization of groundnut (*Arachis hypogaea* L.) composite collection. In: Project abstracts, GCP annual meeting, Bangkok, Thailand, pp 51–52

Upadhyaya HD, Dwivedi SL, Varshney RK, Hoisington DA, Gowda CLL (2008b) Using genetic and genomic resources to broaden the genetic base of cultivated groundnut. In: Third International Conference of the Peanut Research Community on Advances in *Arachis* through Genomics and Biotechnology. Zhengzhou, China, p 6 (abstract)

Valls JFM, Rao VR, Simpson CE, Krapovickas A (1985) Current status of collection and conservation of South American groundnut germplasm with emphasis on wild species of *Arachis*. In: Moss JP (ed.) Proc. Intern. Workshop on Cytogenetics of *Arachis.* October 31–November 2, 1983. ICRISAT, Patancheru, A. P. India, pp 15–35

Varshney RK, Mahendar T, Aruna R, Nigam SN, Neelima K, Vadez V, Hoisington DA (2009) High level of natural variation in a groundnut (*Arachis hypogaea* L.) germplasm collection assayed by selected informative SSR markers. Plant Breed. 128:486–494

Xi X-Y (1991) Development and structure of pollen and embryo sac in peanut (*Arachis hypogaea* L.). Bot Gaz 152:164–172

Zamski E, Ziv M (1976) Pod formation and its geotropic orientation in the peanut, *Arachis hypogaea* L., in relation to light and mechanical stimulus. Ann Bot 40:631–636

Ziv M (1981) Photomorphogenesis of the gynophore, pod and embryo in peanut, A*rachis hypogaea* L. Ann Bot 48:353–359

# Cytological Features of Peanut Genome

4

Guillermo Seijo, Sergio S. Samoluk, Alejandra M. Ortiz, María C. Silvestri, Laura Chalup, Germán Robledo and Graciela I. Lavia

**Abstract**

This chapter aims to update the chromosomal features evaluated by classical and molecular cytogenetic techniques. Karyotype variability detected within and among species was very useful to unravel the taxonomy of the genus and to establish relationships among species. This chapter includes analyses of chromosome morphology, heterochromatin, rDNA loci, as well as dispersed and clustered repetitive sequences. A critical review of the genome sizes of *Arachis* species is also provided. The usefulness of chromosome data is presented in three examples. The first one deals with the origin of the cultivated peanut. Molecular cytogenetics evidenced that the varieties of *A. hypogaea* may have had a single genetic origin, that *A. monticola* is a direct tetraploid ancestor of peanut, and that *A. duranensis* (A genome) and *A. ipaënsis* (B genome) are the diploid progenitors of the AABB tetraploids. The second one pointed to the analysis of the origin of the rhizomatous tetraploids and their relation to the unique diploid species (*A. burkartii*) of section Rhizomatosae. The cytogenetic data suggest that *A. burkartii* has to be discarded as a genome donor of the tetraploids, and that the latter may have had independent origins involving different species. The third one concerns the species of section Arachis, and how the chromosome data aided in the establishment of the genome groups (A, B, D, F, G, and K).

G. Seijo · S.S. Samoluk · A.M. Ortiz · M.C. Silvestri · L. Chalup · G. Robledo · G.I. Lavia (✉)
Instituto de Botánica del Nordeste (CONICET-UNNE), Sargento Cabral 2131, C.C. 209, 3400 Corrientes, Argentina
e-mail: graciela.lavia@yahoo.com.ar

G. Seijo · A.M. Ortiz · G. Robledo · G.I. Lavia
Facultad de Ciencias Exactas y Naturales y Agrimensura, Universidad Nacional del Nordeste, Av. Libertad 5460, 3400 Corrientes, Argentina

## 4.1 Introduction

The genus *Arachis* is composed of 81 species (Krapovickas and Gregory 1994; Valls and Simpson 2005; Valls et al. 2013; Santana and Valls 2015) distributed within a large region of South America, which extends from the eastern foothills of the Andes Mountains in Bolivia and



R.K. Varshney et al. (eds.), *The Peanut Genome*, Compendium of Plant Genomes, DOI 10.1007/978-3-319-63935-2_4

northern Argentina to the Atlantic coast of Brazil, and from the southern limit of the Amazonian rainforest toward the northern coast of La Plata River in Uruguay. Based on morphology, cross-compatibility, viability of the hybrids, geographic distribution, and cytogenetics, the *Arachis* species have been arranged in nine taxonomic sections: Trierectoides, Erectoides, Extranervosae, Triseminatae, Heteranthae, Caulorrhizae, Procumbentes, Rhizomatosae, and Arachis (Krapovickas and Gregory 1994; Fernández and Krapovickas 1994; Lavia 1996; Valls and Simpson 2005). Cross-compatibility, karyotypic, and meiotic analyses also allowed the identification and description of six different genomes within the section *Arachis*, namely, A, B, D, F, G, and K (Smartt et al. 1978; Stalker 1991; Robledo and Seijo 2008, 2010; Robledo et al. 2009; Silvestri et al. 2015). The genomic constitution of the remaining species of the genus, in the absence of comprehensive cytogenetic and molecular analyses, is less precise and has been traditionally assigned to different genomes on the basis of the subgeneric divisions, that is, Am (Heteranthae), C (Caulorrhizae), E (Trierectoides, Erectoides and Procumbentes), Ex (Extranervosae), T (Triseminatae), and R (Rhizomatosae) (Smartt and Stalker 1982).

Classical and modern molecular cytogenetics revealed a huge variability within and among species of different sections. These studies provided important information about the complexity of the peanut genome and were very useful to unravel the taxonomy of the genus and to establish relationships among the wild species and between them and the cultivated peanut. Here we present an update of the cytological information on *Arachis* species and some examples in which the use of chromosome markers was decisive to understand critical and long-lasting problems in the genus.

## 4.2 General Chromosome Features

The available information evidences that the genus *Arachis* is dibasic, with a great predominance of the species with $x = 10$, and only four species with $x = 9$ (Fernández and Krapovickas 1994; Lavia 1996, 1998; Peñaloza et al. 1996; Peñaloza and Valls 2005). Most of the species of the genus are diploids, except the tetraploids *A. hypogaea* and *A. monticola* of section Arachis, three tetraploid species of section Rhizomatosae, and one triploid accession of *A. pintoi* of section Caulorrhizae.

Chromosome identification in peanut began with the analyses done by Husted (1933, 1936), who distinguished one pair of chromosomes that borne an extended secondary constriction (B chromosome pair, SAT chromosomes after Fernández and Krapovickas 1994) and another one that was conspicuously smaller than any of the others of the complement (A chromosome pair, A9 after Robledo et al. 2009). The karyotypes of peanut and *A. monticola* are highly symmetric with a predominance of metacentric chromosomes. The most common karyotype formulae are 38 m + 2 sm and 36 m + 4 sm. Except the A9 pair, the other chromosomes of the karyotype are very similar in size (1.88 μm on average), whereby the chromosomes of the A and B genomes are indistinguishable by classical techniques (reviewed in Fernández and Krapovickas 1994; Lavia and Fernández 2004). Cytogenetic studies covering the six taxonomic varieties of peanut did not reveal major karyotype differences among the subspecies *hypogaea* and *fastigiata* (Lavia and Fernández 2004). Only one (very rarely two) pair of SAT chromosomes has been usually distinguished in all the varieties and in *A. monticola*. However, different types of SAT chromosomes, types 3, 5, and 6, according to the classification of Fernández and Krapovickas (1994), have been observed among the landraces (Lavia and Fernández 2004).

The three tetraploid species of section *Rhizomatosae* have their complements mainly constituted by metacentric (m) chromosomes of similar size, with one pair of SAT chromosomes type 3. Among them, *A. pseudovillosa* has a karyotype formula of 40 m, whereas *A. nitida* and both varieties of *A. glabrata* have 38 m + 2 sm (Fernández and Krapovickas 1994; Peñaloza and Valls 2005; Ortiz et al. 2014).

Among the diploid species, only those from the sections Arachis, Caulorrhizae, Heteranthae,

and Rhizomatosae were comprehensively studied. Within species with $2n = 20$ of section Arachis, 15 are characterized by the presence of the A9 chromosomes and have symmetric karyotypes mainly composed of metacentric chromosomes (Fernández and Krapovickas 1994; Lavia 1996, 2000; Peñaloza and Valls 2005; Robledo et al. 2009). The most frequent karyotype formula is 18 m + 2 sm. The remaining species with $2n = 20$ (genomes B, F, and K, after Robledo and Seijo 2010) have symmetric karyotypes, without any small chromosome (Smartt et al. 1978; Smartt and Stalker 1982; Fernández and Krapovickas 1994; Lavia 1996; Peñaloza and Valls 2005). The most frequent karyotype formula is 20 m, but one to four pairs of submetacentric chromosomes were observed in different species (Fernández and Krapovickas 1994; Robledo and Seijo 2010). *Arachis glandulifera* (D genome) is the only species ($2n = 20$) that has an asymmetric karyotype with a karyotype formula of 8 m + 4 sm + 8 st (Stalker 1991; Robledo and Seijo 2008). Diploid species with $2n = 18$ (G genome, after Silvestri et al. 2015) have symmetric karyotypes with all the chromosomes of similar size (Lavia 1996, 1998; Peñaloza et al. 1996). Almost all the species of the section Arachis have only one pair of secondary constrictions localized on the long arms of pair 10. The exceptions are *A. krapovickasii* and *A. gregoryi*, which have two pairs of secondary constrictions (Robledo and Seijo 2008, 2010; Robledo et al. 2009).

The species of section *Heteranthae* have a karyotype formula of 18 m + 2 sm and SAT chromosomes type 2 (*A. dardani*, *A. pusilla*, and *A. interrupta*), or 16 m + 4 sm and with SAT chromosomes type 10 (*A. sylvestris* and *A. giacomettii*) (Silva et al. 2010). The karyotype formula of the *Caulorrhizae* species is 18 m + 2 sm, but *A. pintoi* has a pair of SAT chromosomes type 2, while *A. repens* has type 3 (Pucciariello et al. 2013). The formula of *A. burkartii* of section *Rhizomatosae* is 20 m with SAT chromosomes type 8 (Ortiz et al. 2013).

## 4.3 Heterochromatin

Chromatin has several classes of proteins complexed with its DNA which are responsible for the major levels of its compaction. Of these levels of chromatin compaction, one of the most conspicuous is the fraction that remains highly condensed throughout the cell cycle, usually defined as constitutive heterochromatin. Although heterochromatin is largely the gene-poor fraction, nowadays it is considered that it plays important roles in the modulation of gene expression, in chromosome structure, and in speciation and evolution of eukaryotes (reviewed in Grewal and Moazed 2003). Accordingly, the analyses of heterochromatin in *Arachis* species revealed extremely different patterns among species and provided many chromosome and genome markers (Seijo et al. 2004; Robledo and Seijo 2008, 2010; Robledo et al. 2009; Custodio et al. 2013; Silvestri et al. 2015).

In the complements of peanut and *A. monticola*, half of the chromosomes (those of the A genome) have centromeric $DAPI^+$ bands (AT-rich), while the remainder (those of the B genome) lack detectable centromeric heterochromatin. The bands of the A complement differ in size, with the most conspicuous being those borne by the A9 pair. All the *A. hypogaea* varieties and *A. monticola* have a similar distribution and amount of heterochromatin, which accounts for about 7% of the karyotype length (Seijo et al. 2004).

Among the diploid species of section Arachis, one of the most variable characters is the amount and distribution of heterochromatin. The A genome species are characterized by having conspicuous $DAPI^+$ centromeric bands in all (or nine) of the chromosome pairs and a total amount of heterochromatin of around 12% of the karyotype length. The size of the bands is around 10% of the chromosome length, except in the A9 and A10 pairs, in which the band size ranged from 25 to 46% and from 14.50 to 20.12%, respectively, among species (Robledo et al. 2009).

The pattern of $DAPI^+$ heterochromatin in the K genome species is characterized by conspicuous centromeric bands in nine chromosome pairs (Robledo and Seijo 2010). The total amount of centromeric heterochromatin per haploid complement is around 12%. Pairs K9 and K10 have the largest bands (16–20%), and the remaining chromosomes have the smallest ones (around 10%) in relation to the chromosome length.

The karyotypes of the F genome species characteristically had small and faint $DAPI^+$ centromeric bands in only seven or eight chromosome pairs (Robledo and Seijo 2010). All the bands had a similar size (around 8% of the chromosome length), and the total amount of heterochromatin per haploid complement was almost half (around 6.5%) of that observed in the species of the A and K genomes.

The B genome species had karyotypes devoid of detectable centromeric bands. However, few species had one small interstitial or distal band in the short arms of pair B3, which usually covers less than 10% of the chromosome length (Robledo and Seijo 2010).

The three $x = 9$ species of section *Arachis* have centromeric $DAPI^+$ bands with the same brightness, position, and size in all chromosome pairs, except *A. palustris*, which lacks these bands in pair G5 (Silvestri et al. 2015).

Aside the species of section *Arachis*, karyotypes with centromeric $DAPI^+$ bands in all (or most) of the chromosome pairs are found in all the sections so far analyzed—Erectoides, Heteranthae, Procumbentes, Triseminatae, Caulorrhizae, and Rhizomatosae (Raina and Mukai 1999; Lavia et al. 2011; Pucciariello et al. 2013; Ortiz et al. unpublished). $CMA^+/DAPI^-$ (GC-rich) heterochromatin is restricted to the secondary constrictions of SAT chromosomes in the species of sections Arachis, Caulorrhizae, and Heteranthae studied so far, except *A. dardani* and *A. giacomettii* that present centromeric $CMA^+$ bands in one chromosome pair (pairs 3 and 5, respectively) and *A. pusilla* that has centromeric GC-rich heterochromatin in the centromeric regions of all the chromosomes (Silva et al. 2010).

## 4.4 rDNA Loci

Localization of the 5S and 18S-26S rRNA genes on the chromosomes by fluorescent in situ hybridization (FISH) was initially applied to a small set of species from different sections of *Arachis* (Raina and Mukai 1999) revealing their usefulness for the characterization of the species. Chromosome mapping of these loci was later used to analyze in detail the karyotypes of all the species included in section *Arachis* (Seijo et al. 2004; Robledo and Seijo 2008; Robledo et al. 2009, 2010; Robledo and Seijo 2010; Lavia et al. 2011; Custodio et al. 2013; Silvestri et al. 2015).

Physical mapping of the rDNA loci in the six botanical varieties of *A. hypogaea* and in *A. monticola* revealed two pairs of 5S and five pairs of 18S–26S rDNA sites. In both species, the 5S loci are proximally located in short arms (pairs A3 and B3), while the 18S–26S rDNA loci are proximally (pairs A2, A10, B3, and B10) or subterminally placed (B7). One 5S locus is syntenic with a 18S–26S site in the pair B3. The high degree of homeology detected between *A. monticola* and *A. hypogaea* strongly evidences that they are very closely related taxa. The mapping of the rDNA loci, together with the heterochromatin analysis, provided the first chromosome map for peanut (Seijo et al. 2004).

Regarding diploid species of section Arachis, those within the A genome have only one interstitial (or rarely proximal) 5S rDNA locus located in the A3 pair. However, the number, size, and chromosomal localization of the 18S–26S rDNA loci vary among the species (Robledo et al. 2009). The number of these gene clusters ranges from two to four pairs, and they present variable size. In general, the largest loci are located in pair A10, those of intermediate size in pair A2, while the smallest and faintest signals (in the cases that the species have more than two loci) in the pairs A7 and A4. According to the pattern of rDNA loci and the heterochromatic bands, the A genome species have been further arranged into three karyotype groups (Robledo et al. 2009): Chiquitano, Pantanal, and La Plata River Basin.

*Arachis glandulifera* (D genome) has only one 5S rDNA locus subterminally located in long arms of pair D5 and five 18S–26S rDNA loci located in different positions and arms of the D1, D2, D6, D9, and D10 pairs. The markers so far identified were enough to the precise identification of all the chromosome pairs of the karyotype and to the construction of the first wholly resolved idiogram for an *Arachis* species (Robledo and Seijo 2008).

All the other species of section Arachis have one 5S rDNA locus localized in proximal or interstitial position on the short arms of the metacentric pair 3. Exceptions are the species of the K genome that have two additional pairs of loci located in the pairs K8 and K10. The number of 18S–26S rDNA loci ranges from two (in *A. gregoryi* and *A. trinitensis*) to four (in *A. magna* and *A. valida*). Most of them are located in pericentromeric or interstitial position on the long arms. In most species, the largest and brightest 18S–26S rDNA loci correspond to the clusters located in the secondary constrictions of the SAT chromosomes (pair 10), while the remaining ones are small and pale. In some species, one 18S–26S rDNA and one 5S rDNA loci are localized on the same chromosome. In the species of the K genome, these loci co-localize on the long arm of pair K10, while in *A. benensis* and *A. magna*, they co-localize on the short arm of pairs F3 and B3, respectively. In *A. ipaënsis*, one 18S–26S rDNA and one 5S rDNA loci map to pair B3, but on different arms (Robledo and Seijo 2010).

The three species with $x = 9$ have only one 18S–26S rDNA site in the proximal region of the long arm of the G9 pair (SAT chromosomes) and one 5S rDNA site in the short arm of the pair G6 (Silvestri et al. 2015).

## 4.5 Repetitive Sequences

Plant genomes are composed of single-copy sequences, with one or few copies (gene sequences), and repetitive sequences, with a higher copy number. The latter can be found as dispersed repetitive (transposons) or tandemly repetitive (satellite DNA) sequences (Schmidt and Heslop-Harrison 1998). One of the most important features of the repetitive genome component is its rapid evolution both at the sequence level and genome representation (Schmidt and Heslop-Harrison 1998). For this reason, the analysis of this fraction is a useful tool for the study of evolutionary relationships between plant species (Dechyeva et al. 2003; Navajas- Pérez et al. 2009; Nielen et al. 2010, 2012; Samoluk et al. 2015a). Moreover, many of these sequences when probed onto chromosomes provide conspicuous markers for genome characterization, for the establishment of homeologies and for the construction of chromosome maps (Seijo et al. 2004; Robledo et al. 2009; Robledo and Seijo 2010; Zhang et al. 2016).

## 4.6 Dispersed Sequences

Studies based on retrotransposons at a genomic scale are scarce in *Arachis* and have been focused on peanut and its wild diploid progenitors (*A. ipaënsis* and *A. duranensis*). To date, Ty3-gypsy (Nielen et al. 2010) and Ty1-copia (Nielen et al. 2012) LTR retrotransposons have been characterized and quantified in these three species. A significant differential representation of Ty3-gypsy retrotransposons, but not of Ty1-copia retrotransposons, was described in the two diploid species. The element, named FIDEL (Fairly long Inter-Dispersed Euchromatic LTR retrotransposon), is more frequent in the A than in the B genome, with copy numbers of about 3000 (± 950, *A. duranensis*), 820 (± 480, *A. ipaënsis*), and 3900 (± 1500, *A. hypogaea*) per haploid genome. Phylogenetic analysis of reverse transcriptase sequences showed the distinct evolution of FIDEL in the diploid species. Fluorescent in situ hybridization revealed a disperse distribution in the euchromatin and absence from centromeres, telomeric, and the nucleolar organizer regions. Distribution of FIDEL onto the chromosomes reflects almost the pattern of GISH using genomic probes of the diploid progenitors onto the chromosomes of peanut (Seijo et al. 2007).

By contrast, the Ty1-copia retrotransposon from the Bianca lineage (named Matita) is a moderate copy number element (Nielen et al. 2012). This element is almost equally represented in the A and B genomes in relatively low copy. FISH experiments showed that Matita is mainly located in the distal regions of the chromosome arms and its chromosome-specific hybridization pattern aided in the identification of some individual chromosomes. By probing BAC libraries with overgos probes from Matita, it was demonstrated that this element is not randomly distributed in the genome, but exhibits a significant tendency of being more abundant near resistance gene homologues than near single-copy genes.

A more recent study comparing 1.26 Mb of homeologous A and B genomes BAC clones evidenced the existence of a diverse group of complete and truncated copies of the LTR retrotransposons fraction that covered more than 40% of the sequences analyzed (Bertioli et al. 2013). BAC-FISH using 27 *A. duranensis* BAC clones as probes gave dispersed and repetitive DNA characteristic signals, predominantly in interstitial regions of the peanut A chromosomes. In general, the sequences of 14 BAC clones revealed that a substantial proportion of the highly repetitive component of the peanut A genome is represented for relatively few LTR retrotransposons and their truncated copies of LTRs.

Non-LTR retroelements are generally less abundant than LTR retroelements in the plant genomes (Noma et al. 1999; Alix and Heslop-Harrison 2004; Hawkins et al. 2006), and *Arachis* genomes are not the exception. The available data reveal that the genome content of these elements is less than 8% in *A. duranensis* (Chen et al. 2016) and approximately 12% in *A. ipaënsis* (Bertioli et al. 2013). However, the genome representation of these retroelements is relatively high when compared to the amount of LINEs present in other plant species (Samoluk et al. 2015b; Bertioli et al. 2016). Recently, the diversity, the chromosome distribution, and the genome representation of a LINE family belonging to the L1 clade were analyzed in six genomes and karyotype groups of section *Arachis* (Samoluk et al. 2015b). The phylogenetic analysis based on the reverse transcriptase of these elements showed that the lineages are distributed independently of the genomes or karyotype groups. FISH experiments revealed a dispersed pattern with hybridization signals mainly located on the euchromatin of interstitial and distal regions of most chromosome arms in all the genome types analyzed in that study. In agreement with the results obtained by Bertioli et al. (2016), the genome abundance of this kind of retroelements was higher in *A. ipaënsis* than in *A. duranensis*.

The sum of available data evidences that retroelements have a dispersed pattern in all the genomes analyzed so far, although with different representation among them. Though some of these elements may be used for the identification of different chromosome complements in hybrids and allopolyploids, their usefulness as cytogenetic markers for individual chromosome identification is limited.

In contrast to the high genome abundance of retroelements, DNA transposons constitute about 5–10% of the genome of *Arachis* species (Bertioli et al. 2016; Chen et al. 2016). However, these elements have been little explored in *Arachis*, and there is no available data on their chromosome distribution. In this sense, there are some reports about a miniature inverted-repeat transposable element, the AhMITE1 element (Shirasawa et al. 2012; Gowda et al. 2010, 2011). This element is present in high copy number in the genomes of *A. hypogaea*, *A. magna*, and *A. monticola*, but not in *A. duranensis* (Shirasawa et al. 2012). In addition, the study of AhMITE1 transpositional activity among different allotetraploid *Arachis* species showed that most of the *A. hypogaea* subsp. *fastigiata* types carry an AhMITE1 insertion at the FST1-linked site, whereas the wild allotetraploid *A. monticola* and *A. hypogaea* subsp. *hypogaea* types missed the AhMITE1 element at that site (Gowda et al. 2011). This finding supports that the *A. hypogaea* subsp. *hypogaea* is closer to the wild allotetraploid than the subsp. *fastigiata* (Paik-Ro et al. 1992; Singh et al. 1993; He and Prakash 2001).

Moreover, it proposed that AhMITE1 transposition could have been of major importance in the origin of *A. hypogaea* subsp. *fastigiata* (Gowda et al. 2011).

## 4.7 Clustered Sequences

Satellite DNA constitutes a significant portion of eukaryote genomes. It is formed by repetitive units of variable length (140–180 bp or 300–360 bp) tandemly arranged in blocks of up to 100 Mpb (Charlesworth et al. 1994; Schmidt and Heslop-Harrison 1998; Plohl et al. 2008). These sequences usually show particular chromosome locations, being a major component of the centromeric (Hudakova et al. 2001; Gindullis et al. 2001; Urdampilleta et al. 2009), telomeric (Pich et al. 1996; Macas et al. 2000), and less frequently, interstitial heterochromatin (Mukai et al. 1992). Therefore, they have become a useful tool to study the karyotype evolution in different groups of species (Lanfredi et al. 2001; Slamovits et al. 2001).

The analysis of a satellite sequence named ATR-2 in seven diploid species ($x = 10$) representative of different genomes and karyotype groups revealed an infraspecific and interspecific conservation of these sequences, with a low spreading of new monomeric variants in the analyzed species (Samoluk et al. unpublished). However, the quantitative analyses revealed differences in the abundance of this satellite DNA among them, according to the predictions of the "library hypothesis" (Fry and Salser 1977). FISH analyses revealed that ATR-2 is exclusively distributed at the $DAPI^+$ centromeric heterochromatin; however, it may not be the only sequence that conforms this genomic fraction. Despite the sequence conservation of ATR-2, the variable representation of this satDNA suggests that it was actively involved in the remodeling the heterochromatic patterns of the diploid *Arachis* species (Samoluk et al. unpublished).

Another major repetitive DNA sequences were cloned and analyzed from *A. hypogaea* Cot-1 DNA in order to identify new genome- and chromosome-specific markers (Zhang et al. 2012). In particular, a satellite DNA sequence of 115 bp was found mainly distributed in pericentromeric regions on most of the B genome chromosomes of the allotetraploid (Zhang et al. 2012). Recently, it was developed an FISH-based karyotyping system using a set of new and previously reported chromosome markers, which allowed the identification of almost all chromosomes and the construction of karyotypes in cultivated peanut and its two putative progenitors (Zhang et al. 2016).

## 4.8 Genome Size

Genome size is a useful descriptor for characterization of plant genetic resources (Ozias Akins and Jarret 1994; Rayburn et al. 1997; Hendrix and Stewart 2005; Loureiro et al. 2007). Until recently, nuclear DNA contents were studied in a very limited number of *Arachis* species, and the available genome size estimations were controversial. Most of these determinations were made by the Feulgen densitometry method (Dhillon et al. 1980; Resslar et al. 1981; Singh et al. 1996; Lavia and Fernández 2008). However, measurements in *A. duranensis* and *A. hypogaea* by flow cytometry indicated that the data obtained by Feulgen densitometry [except in Dhillon et al. (1980) for *A. hypogaea*] overestimated the genome size of *Arachis* species by twofold (Temsch and Greilhuber 2000, 2001).

In a recent study, 26 diploid species of the section *Arachis* were analyzed by flow cytometry, and the 2C values ranged from 2.55 to 3.22 pg (Samoluk et al. 2015b). The annual species belonging to different genomes (Robledo et al. 2009; Robledo and Seijo 2010) tend to have different genome sizes. However, the 2C values of the perennial species (all with A genome) are distributed almost continuously along the whole range of genome sizes (2.55–3.22 pg) of the annual species. The comparison of 2C values with karyotype parameters suggests that changes in DNA content have been proportionally distributed among the chromosome arms, and that the heterochromatic fraction is not directly involved in those changes. Within the A genome,

the annual species has lower DNA content than the perennial ones, which is in accordance with the predictions of the nucleotype hypothesis (Bennet 1982). However, the lack of significant relationships with geoclimatic variables suggests that there are many intrinsic factors determining particular ecological roles of the DNA content in different lineages of section Arachis. A critical analysis of the DNA content of other species of *Arachis* is still needed to address the direction of the genome change during the evolution of the genus as a whole.

Measurements of the genome size of the AABB species of *Arachis* showed that they are among the few allopolyploids in which their genome size are about the sum of those of their diploid progenitors (Samoluk et al. 2015b), like in tobacco (Leitch and Bennett 2004), *Hordeum* (Jakob et al. 2004), and AD *Gossypium* species (Wendel et al. 2002). The genome sizes estimated for *A. monticola* (5.70 pg) and for the cultivated peanut (5.60 pg) were *on par* with the expected 2C value estimated from the sum of the genome sizes of their parental species (*A. duranensis*, 2C = 2.55 pg, and *A. ipaënsis*, 2C = 3.19 pg) (Samoluk et al. 2015b). The constancy in the Cx values suggests that the hybridization and chromosome doubling events that occurred during the origin of the cultivated peanut have not resulted in significant changes in genome size.

## 4.9 Usefulness of Chromosome Markers

The use of chromosome markers in *Arachis* species came to complete and extends a large number of the taxonomic, classical cytogenetic, cross hybridization, molecular marker, and geographical studies. Here we describe two cases in which the use of these chromosome markers contribute to shed light on long debates among researchers: one is about the origin of peanut and the other deals with the genome characterization of the *Arachis* species.

### 4.9.1 Inferences on Peanut Origin

The origin of peanut has been a matter of study for several decades and has long been assessed from different points of view. The identification of one pair of small chromosomes (A9 pair after Robledo et al. 2009), and one pair of SAT chromosomes (A10 pair after Fernández and Krapovickas 1994), instead of four chromosomes of each type led to the proposal that the peanut is an allotetraploid species with $2n = 4x = 40$ (Husted 1933, 1936), and with an AABB genome constitution (Smartt et al. 1978). This has been confirmed by studies on interspecific hybridization among the cultivated peanut and different wild diploid species (Smartt and Gregory 1967; Stalker and Wynne 1979; Singh 1986), and by modern cytogenetic techniques (Seijo et al. 2004, 2007).

However, the diploid species that were involved in the origin of cultivated peanut were until recently under debate. Before the development of chromosome markers, more than eight wild diploid species having different genome types were considered involved in the origin of peanut (reviewed in Singh and Smartt 1998; Seijo et al. 2007; Grabiele et al. 2012). Studies based on molecular markers showed that several species of the A genome could be considered as the most probable ancestor of peanut. Restriction fragment length polymorphism (RFLP) revealed *A. duranensis* as the most probable candidate (Kochert et al. 1991, 1996), whereas randomly amplified polymorphic DNA (RAPD) and inter-simple sequence repeat (ISSR) analyses showed *A. villosa* as the best candidate (Raina et al. 2001). On the other hand, PCR amplified fragment length polymorphism (AFLP) data have shown that at least three diploid species of the A genome were closely related to the cultigen (Milla et al. 2005). Similarly, microsatellite markers have revealed that, although *A. duranensis* is the most closely related to the cultigen, a small group of other species having the A genome could also be genome donors of peanut (Moretzsohn et al. 2004).

Classical cytogenetics aided in the identification of the diploid progenitors of peanuts. However, only when chromosome banding and molecular cytogenetic techniques were applied massively to *Arachis* species by determining the patterns of heterochromatin and the number (Raina and Mukai 1999) and position (Seijo et al. 2004; Robledo et al. 2009; Robledo and Seijo 2010; Custodio et al. 2013) of the rRNA genes by FISH, the wild diploid progenitors were more precisely identified. The analysis of the rDNA loci distribution showed that the species of the A genome included in the La Plata River Basin group are the most related to the A genome of the tetraploids, and that *A. ipaënsis* is the most probable B genome donor. Subsequently, double GISH experiments using genomic DNA of the diploid *Arachis* species ($2n = 2x = 20$) identified *A. duranensis* as the A genome donor of *A. hypogaea* (Seijo et al. 2007).

*Arachis monticola* is the only wild allotetraploid within section *Arachis*, and it is currently known only from three very close localities in NW of Argentina. In all the dendrograms constructed using molecular markers, the cultivated peanut and *A. monticola* group together, with very low or no genetic distance (Gimenez et al. 2002; Grabiele et al. 2012; Moretzohn et al. 2013). They are known to be interfertile, with no apparent sterility in the $F_1$ hybrids (Kirti et al. 1983). Therefore, *A. monticola* is regarded by different authors either as the direct progenitor of the peanut or as an introgressive derivative between the peanut and wild species (see Stalker and Moss 1987; Grabiele et al. 2012). However, if *A. hypogaea* (AABB) could cross with any diploid species of the A or B genomes through reduced gametes (n), the result would be an infertile or very poorly fertile triploid with AAB or ABB genome constitutions, respectively. Alternatively, if crosses between *A. hypogaea* and diploid species of the A or B genomes had occurred via unreduced gametes ($2n$) of the wild species, then the hybrids would have AAAB or ABBB genome constitutions, respectively. Neither of these scenarios would have given rise to an allotetraploid with the genome constitution compatible with that of *A. monticola* (AABB) (Seijo et al. 2004, 2007; Grabiele et al. 2012). This rationale suggested that the hypothesis that considers *A. monticola* as an introgressive derivative has to be discarded in the light of the molecular cytogenetic data, since the wild tetraploid has exactly the same chromosome complements to that of *A. hypogaea.*

The fact that the amphidiploid that resulted from the artificial resynthesis from *A. ipaënsis* and *A. duranensis* (Fávero et al. 2006) is morphologically very similar to *A. monticola*, and that it can hybridize with all the varieties of the cultigen producing fertile offspring, supports the hypothesis that considers *A. monticola* as the direct progenitor of *A. hypogaea*. Moreover, the ability to persist in natural populations (unlike the cultivated peanut) and the wild type structure of its fruits (wherein each seed has its own shell separated by an isthmus) support the maintenance of *A. monticola* as a separate taxonomic species.

The identical patterns of molecular cytogenetic markers and genomic hybridization (GISH) detected in all the botanical varieties of the cultigen suggests that the same wild species participated in their origin. Moreover, this finding implies that all the presently known varieties of peanut arose from a single, unique allotetraploid plant population (Seijo et al. 2004; Seijo et al. 2007). The common ancestry of all infraspecific taxa of *A. hypogaea* is supported by the low genetic variability so far detected with most molecular markers in the cultivated peanut (Halward et al. 1991; Kochert et al. 1996; Herselman 2003; Grabiele et al. 2012; Moretzohn et al. 2013).

In this scenario, after the origin of the wild allotetraploid (which probably had larger seeds than any of the progenitors as a result of the gigas effects in polyploids, like in *A. monticola*), *A. hypogaea* may have arisen through domestication. Therefore, the large morphological, ecological, phenological, and chemical variability present in the many landraces of peanut (Krapovickas et al. 2009, 2013) would have mainly resulted from particular selective

pressures undergone in different agroecological environments (Krapovickas and Gregory 1994; Grabiele et al. 2012).

### 4.9.2 Origin of Other Tetraploid *Arachis* Species

Section *Rhizomatosae* is currently defined exclusively on morphological features, mainly because all (four) the species have rhizomes and can be asexually propagated. Among this group of taxa, *A. burkartii* is the only diploid with $2n = 2x = 20$, while the others, *A. pseudovillosa, A. nitida*, and *A. glabrata*, are tetraploids with $2n = 4x = 40$ (Gregory et al. 1973; Fernández and Krapovickas 1994; Peñaloza and Valls 2005). Interestingly, the tetraploid species are distributed from the Mato Groso State in Brazil to the North of Argentina, while the diploid species live from the North of Argentina to the North of Uruguay. Thus, the species with different ploidy level of section *Rhizomatosae* only overlaps in a very narrow stretch in the NE of Argentina (Krapovickas and Gregory 1994; Valls and Simpson 2005).

If this section were monophyletic, it would be expected that *A. burkartii* was one of the genome donors of the rhizomatous tetraploids. However, classical cytogenetic and molecular data argued the monophyly of the section and the nature of the polyploids. The meiotic analysis of *A. nitida* revealed that the chromosomes are arranged in 20II in most (65%) of the cells, although up to four multivalents (trivalents and quadrivalents) in low frequencies were observed (Ortiz et al. 2011). Moreover, it has one 5S rDNA locus, two 18S–26S rDNA loci, and a similar pattern of centromeric heterochromatic bands in the four chromosome sets. However, the presence of one 18S–26S rDNA sites in only two of the four chromosomes of pair 2 of *A. nitida* suggests that it has two different chromosome sets. Thus, *A. nitida* may be considered a segmental allopolyploid (Ortiz et al. unpublished).

Controversial hypotheses have been put forward for the nature of *A. glabrata*. Some authors proposed that this species is a true autopolyploid (Singh and Simpson 1994; Ortiz et al. 2011), while others suggested that it may be an allopolyploid with the EERR genome constitution (Gregory and Gregory 1979; Bechara et al. 2010). Meiotic analyses of this species showed that the frequency of bivalents differs significatively (from 20 to 81%) among accessions (Ortiz et al. 2011). Also, the number of trivalents (from 1 to 3III) and quadrivalents (from 1 to 8IV) was very variable among them. However, based on the fact that the four chromosome sets have $DAPI^+$ centromeric bands in all the chromosomes and a similar pattern of the 5S and 18S–26S rDNA loci, together with the detection of up to eight quadrivalents in meiotic cells, it was suggested that this species may be either a true autopolyploids or, less probably, a segmental allopolyploid with different degrees of diploidization (Ortiz et al. unpublished).

Concerning *A. pseudovillosa*, since meiotic behavior studies could not be performed yet, the polyploidy nature is still under study (Ortiz et al. unpublished). In this sense, the presence of one 5S and one 18-26S rDNA sites, and the similar distribution pattern of CMA–DAPI bands in the four chromosome sets suggest that *A. pseudovillosa* may be an autopolyploid. However, the presence of an extra 18S–26S rDNA locus in only one chromosome set may have arisen de novo by transposition (or other genomic mechanism) after polyploidization, or may be interpreted as evidence for a segmental allopolyploid origin by hybridization between two closely related species which differ in the numbers of this rDNA cluster genes (Ortiz et al. unpublished).

Molecular markers strongly support the conclusion obtained from the cytogenetic data about the relationship among the rhizomatous species. RAPD (Nobile et al. 2004) and SSR (Angelici et al. 2008) analyses that included the four species of section *Rhizomatosae* showed the clustering of the tetraploid species in a single group, distant from *A. burkartii*. Further, the AFLP analysis including representatives of seven different sections revealed a close association of *A. glabrata* with *A. major* and *A. paraguariensis* (sect. *Erectoides*), while *A. burkartii* was associated with the two species of the section

*Caulorrhizae* (Gimenez et al. 2002). In addition, molecular phylogenies based on chloroplast and nuclear DNA sequences (Bechara et al. 2010; Friend et al. 2010), which only included *A. burkartii* and *A. glabrata*, have shown that the diploid taxon is found in an individual and isolated clade, while *A. glabrata* grouped in a distant clade with members of sections *Erectoides* and *Procumbentes.*

Regarding the genomic constitution of rhizomatous tetraploid species, the cytogenetic evidences suggest that the three species may have at least one common diploid ancestor (Ortiz et al. unpublished). In this sense, all the species of the section *Rhizomatosae* (2*x* and 4*x*) have been traditionally assigned to the R genome (Smartt and Stalker 1982) assuming that the section was monophyletic and the diploid rhizomatous species, *A. burkartii,* was the natural ancestor of the tetraploids. However, the fact that *A. burkartii* has a karyotype formula of 20 m with a SAT chromosome type 8, lacks $DAPI^+$ heterochromatic bands, and presents four pairs of 18S–26S rDNA loci (located on three chromosome pairs) and one pair of interstitial 5S rDNA sites (that co-localized with of 18S–26S rDNA loci on the smallest chromosome pair) evidenced that this diploid species is unlikely the genome donor of the rhizomatous tetraploids. Instead, the complements of these tetraploids showed high homology with those of the *Erectoides* and *Procumbentes* sections. Thus, the tetraploid species should be excluded from the R-genome, which might remain exclusively for *A. burkartii* (Ortiz et al. unpublished).

The aforementioned cytogenetic data do not support the origin of rhizomatous tetraploids from *A. burkartii* and suggest that the section *Rhizomatosae* is not monophyletic (Nóbile et al. 2004; Angelici et al. 2008; Bechara et al. 2010; Friend et al. 2010; Ortiz et al. unpublished).

### 4.9.3 Genome Arrangement of Section *Arachis*

Diploid species of section *Arachis* with $2n = 20$ and symmetric karyotypes have been traditionally arranged first into two different genome groups (A and non-A genomes) on the basis of the presence of the small chromosomes A9 (first observed in *A. hypogaea*) in their karyotypes. These two groups of species show strong reproductive isolation manifested by low hybrid production, and low chromosome pairing and low pollen stainability in the hybrids (Gregory and Gregory 1979; Stalker et al. 1991; Krapovickas and Gregory 1994; Tallury et al. 2005). The only species with $2n = 20$ and asymmetric karyotype (*A. glandulifera*) was assigned to a different genome, the D genome (Stalker 1991).

The development of chromosome markers by molecular cytogenetics revealed a high degree of homogeneity in the karyotypes among the species with A chromosomes. However, variation in number and positions of $DAPI^+$ bands and major 18S–26S rDNA sites among species was used to establish three subgroups of karyotype homeologies (Robledo et al. 2009) considering the fact that closeness of taxa is usually correlated with the similarity of their heterochromatin and rDNA FISH patterns (Hizume et al. 2002; Liu et al. 2003). Since the groups that resulted from the homeology analysis included species that tend to be more closely distributed geographically than those belonging to different groups, they were named using a geographical reference (Robledo et al. 2009). The Chiquitano group comprised the species (*A. cardenasii*, *A. herzogii*, and *A. kempff-mercadoi*) that grow in the southern and western portion of the Chiquitanía biogeographic region in Santa Cruz Department of Bolivia. The Pantanal group includes the species (*A. diogoi*, *A. kuhlmannii*, *A. helodes*, *A. simpsonii*, and *A. stenosperma*) which are distributed in the Pantanal biogeographic region in western Brazil, northern Paraguay, and eastern Bolivia. This group may also include *A. linearifolia*. The La Plata River Basin group corresponds to the species (*A. duranensis*, *A. schininii*, *A. correntina, A. villosa*, and probably *A. microsperma*) that are distributed along the La Plata River Basin (except the region comprising the upper stream of the Paraguay River in the Pantanal).

Most molecular marker studies in a large set of A genome species support the Pantanal group (Kochert et al. 1991; Raina et al. 2001; Milla

et al. 2005; Moretzon et al. 2013). The Chiquitano group is least represented in molecular analyses, but whenever *A. herzogii* and *A. kempff-mercadoi* have been included, they always clustered together (Milla et al. 2005; Tallury et al. 2005). In a recent microsatellite-based phylogenetic analysis, *A. cardenasii* and *A. kempff-mercadoi* were grouped together with the species that belong to the Pantanal group (Moretzohn et al. 2013). However, only two accessions of *A. cardenasii* and none of *A. herzogii* were included in that study. Species belonging to La Plata River Basin group generally clustered together with a few exceptions (Milla et al. 2005; Moretzohn et al. 2013).

The non-A genome species have been segregated into three genomes (B, F, and K) based on the different patterns of chromosome markers (Robledo and Seijo 2010). The K genome includes *A. batizocoi*, *A. cruziana*, and *A. krapovickasii.* These species are characterized by having conspicuous $DAPI^+$ heterochromatic bands in all chromosome pairs except in K7 and three 5S rDNA loci on K2, K4, and K10 pairs. The F genome comprises *A. benensis* and *A. trinitensis*, and their karyotypes have small and faint bands in seven or eight chromosome pairs and only one 5S rDNA locus on F3 pair. The B genome species (after Robledo and Seijo 2010) have karyotypes without pericentromeric $DAPI^+$ heterochromatin and includes *A. ipaënsis*, *A. gregoryi*, *A. magna, A. valida,* and *A. williamsii* (Robledo and Seijo 2010). This group showed the highest homeology with the B genome of *A. hypogaea*. *Arachis glandulifera* has asymmetric karyotype (Stalker 1991; Fernández and Krapovickas 1994), a unique pattern of heterochromatin $DAPI^+$ distribution and 5 pairs of 18S–26S rDNA loci (Robledo and Seijo 2008). On this basis, this species was confirmed as having the D genome as proposed earlier by Stalker (1991).

The proposed genome arrangement is supported by species crossability, pollen stainability, morphological characters, and geographical distribution of the species (Smartt et al. 1978; Stalker 1991; Krapovickas and Gregory 1994; Tallury et al. 2005; Burrow et al. 2009). The position of the D, F, and K genomes with respect to the A and B genomes is controversial. The analysis of chloroplast sequences revealed that these genomes are closely related to the B genome, but the NTS of the 5S rDNA genes and AFLP markers showed that the F genome is indeed close to the B genome, while the D and K are sister groups of the A genome (Tallury et al. 2005; Grabiele et al. 2012). The phylogenetic analysis based on DNA sequence information of three single-copy gene introns was consistent with the current genome classification, since clades contained species with the same genome types (Moretzshon et al. 2013). In this analysis, the species with D, F, and K genome species are close to the A genome species, but the microsatellite analysis done in the same report showed that those genomes types are closer to the B genome than to the A genome. In a more recent analysis using nine intron sequences and GISH, it was shown that the K genome is closer to the B genome than to the A genome of *A. hypogaea* (Leal-Bertioli et al. 2015). The extant inconsistency of available data showed that the D, F, and K genomes are in an intermediate position between the A and the B genomes.

Geographically, the species included within each genome tend to be co-distributed. The species with the F genome are restricted to the lowland savannas of Beni department in Bolivia, while those with the K genome are distributed in the NW of the Chacoan Boreal region. The species with the B genome are more widely distributed in semi-deciduous forests and savannas of the cerrado associated with the Chiquitano planalto and west Pantanal. *Arachis ipaënsis*, known from only one population, was collected from the top of the sand banks of streams in an ecotone between the tucumano-oranence deciduous forest and the chacoan xerophytic forest (Robledo and Seijo 2010).

More recently, the three species with $x = 9$ chromosomes of section *Arachis* were analyzed with the same chromosome markers that were mapped in the $x = 10$ species. These analyses showed that $x = 9$ species present a similar pattern of $DAPI^+$ heterochromatin to that observed in the species assigned to A and K genomes. However, it differs from the former by lack of the

A9 chromosomes and the presence of only one pair of 18S–26S rDNA sites. From the latter, it differs by the lower number of 5S rDNA sites and the lower number of submetacentric chromosomes. Based on all these karyotypic differences together with the reproductive isolation from any other species of section *Arachis*, a new genome type (G genome) was proposed for the $x = 9$ species (Silvestri et al. 2015). Molecular datasets have revealed that the three $x = 9$ species form a compact clade, but different from any other group in section *Arachis* (Bechara et al. 2010; Friend et al. 2010; Moretzshon et al. 2013). Moreover, their relationship with other species of the section is still unclear. In this sense, analysis of microsatellites (Moretzshon et al. 2004, 2013), single-copy gene sequences (Moretzshon et al. 2013), trnT-F cpDNA marker (Tallury et al. 2005), and ITS and 5.8S of the nuclear rDNA (Bechara et al. 2010) suggest that these taxa are more closely related to some non-A genome species, whereas RAPD (Creste et al. 2005) and AFLP (Milla et al. 2005) analyses suggest that they are genetically more similar to A genome species.

## 4.10 Conclusion

In conclusion, in spite of the homogeneity in the chromosome morphology of the species of section *Arachis*, the use of chromosome markers revealed six different karyotype organizations, which agree with the different degrees of reproductive isolation. Whereby, based on these different organizations, the species of section Arachis are currently arranged in six different genomes (A, B, D, F, G, and K).

## References


Alix K, Heslop-Harrison JS (2004) The diversity of retroelements in diploid and allotetraploid *Brassica* species. Plant Mol Biol 54:895–909

Angelici C, Hoshino AA, Nóbile PM et al (2008) Genetic diversity in Rhizomatosae section of the genus *Arachis* (Fabaceae) based on microsatellite markers. Genet Mol Biol 31:79–88

Bechara MD, Moretzsohn MDC, Palmieri DA et al (2010) Phylogenetic relationships in genus *Arachis* based on ITS and 5.8S rDNA sequences. BMC Plant Biol 10:255. doi:10.1186/1471222910255

Bennet MD (1982) Nucleotypic basis of the spatial ordering of chromosomes in eukaryotes and the implications of the order for genome evolution and phenotypic variation. In: Dover GA, Flavel RB (eds) Genome evolution. Academic Press, New York, pp 239–261

Bertioli DJ, Cannon SB, Froenicke L et al (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet 48:438–446

Bertioli DJ, Vidigal BS, Nielen S et al (2013) The repetitive component of the A genome of peanut (*Arachis hypogaea*) and its role in remodelling intergenic sequence space since its evolutionary divergence from the B genome. Ann Bot 112:545–559

Burow MD, Simpson CE, Faries MW et al (2009) Molecular biogeographic study of recently described B- and A-genome *Arachis* species, also providing new insights into the origins of cultivated peanut. Genome 52:107–119

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371:215–220

Chen X, Li H, Pandey MK et al (2016) Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens. PNAS 113:6785–6790

Creste S, Tsai SM, Valls JFM et al (2005) Genetic characterization of Brazilian annual *Arachis* species from sections Arachis and Heteranthae using RAPD markers. Genet Resour Crop Evol 52:1079–1086

Custódio AR, Seijo G, Valls JFM (2013) Characterization of Brazilian accessions of wild *Arachis* species of section Arachis (Fabaceae) using heterochromatin detection and fluorescence in situ hybridization (FISH). Genet Mol Biol 36:364–370

Dechyeva D, Gindullis F, Schmidt T (2003) Divergence of satellite DNA and interspersion of dispersed repeats in the genome of the wild beet *B. procumbens*. Chrom Res 11:3–21

Dhillon SS, Rake AV, Miksche JP (1980) Reassociation kinetics and cytophotometric characterisation of peanut (*Arachis hypogaea* L.) DNA. Plant Physiol 65:1121–1127

Fávero AP, Simpson CE, Valls JFM et al (2006) Study of the evolution of cultivated peanut through crossability studies among *Arachis ipaensis*, *A. duranensis* and *A. hypogaea*. Crop Sci 46:1546–1552

Fernández A, Krapovickas A (1994) Cromosomas y evolución en *Arachis* (Leguminosae). Bonplandia 8:187–220

Friend SA, Quandt D, Tallury SP et al (2010) Species, genomes, and section relationships in the genus *Arachis* (Fabaceae): a molecular phylogeny. Plant Syst Evol 290:185–199

Fry K, Salser W (1977) Nucleotide sequences of HS-a satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. Cell 12:1069–1084

Gimenes MA, Lopes Romero C, Galgaro ML et al (2002) RFLP analysis of genetic variation in species of section Arachis, genus *Arachis* (Leguminosae). Euphytica 123:421–429

Gindullis F, Desel C, Galasso I et al (2001) The large-scale organization of the centromeric region in beta species. Genome Res 11:253–265

Gowda MVC, Bhat RS, Motagi BN et al (2010) Association of high-frequency origin of late leaf spot resistant mutants with *AhMITE1* transposition in peanut. Plant Breed 129:567–569

Gowda MVC, Bhat RS, Sujay V et al (2011) Characterization of *AhMITE1* transposition and its association with the mutational and evolutionary origin of botanical types in peanut (*Arachis* spp.). Plant Syst Evol 291:153–158

Grabiele M, Chalup L, Robledo G et al (2012) Genetic and geographic origin of domesticated peanut as evidenced by 5S rDNA and chloroplast DNA sequences Plant Syst Evol 298:1151–1165

Gregory WC, Gregory MP (1979) Exotic germplasm of *Arachis* L. interspecific hybrids. J Hered 70:185–193

Gregory WC, Gregory MP, Krapovickas A et al (1973) Structure and genetic resources of peanuts. In: Wilson CT (ed) Peanut culture and uses, Amer. Peanut Res Educ, Soc, Stillwater, pp 47–133

Grewal SIS, Moazed D (2003) Heterochromatin and Epigenetic Control of Gene Expression. Science 301:798–802

Halward TM, Stalker HT, Larue EA et al (1991) Genetic variation detectable with molecular markers among unadapted germplasm resources of cultivated peanut and wild species. Genome 34:1013–1020

Hawkins JS, Kim H, Nason JD et al. (2006) Differential lineagespecific amplification of transposable elements is responsible for genome size variation in *Gossypium*. Genome Res 16:1252–1261

He G, Pakrash C (2001) Evaluation of genetic relationships among botanical varieties of cultivated peanut (Arachis hypogaea L.) using AFLP markers. Genet Resour Crop Evol 4:347–352

Hendrix B, Stewart JM (2005) Estimation of the nuclear DNA content of *Gossypium* species. Ann Bot 95:789–797

Herselman L (2003) Genetic variation among Southern African cultivated peanut (*Arachis hypogaea* L.) genotypes as revealed by AFLP analysis. Euphytica 133:319–327

Hizume M, Shibata F, Matsusaki Y et al (2002) Chromosome identification and comparative karyotypic analyses of four *Pinus* species. Theor Appl Genet 105:491–497

Hudakova S, Michalek W, Presting GG et al (2001) Sequence organization of barley centromeres. Nucleic Acids Res 29:5029–5035

Husted L (1933) Cytological studies on the peanut, *Arachis*. I. Chromosome number and morphology. Cytologia (Tokyo) 5:109–117

Husted L (1936) Cytological studies on the peanut, *Arachis*. II. Chromosome number and morphology and behavior, and their application to the problem of the origin of the cultivated forms. Cytologia (Tokyo) 7:396–423

Jakob SS, Meister A, Blattner FR (2004) The considerable genome size variation of *Hordeum* species (Poaceae) is linked to phylogeny, life form, ecology, and speciation rates. Mol Biol Evol 21:860–869

Kirti PB, Bharathi M, Murty UR et al (1983) Chromosome morphology in three diploid species of *Arachis* and its bearing on the genomes of groundnut (*Arachis hypogaea* L.). Cytologia 48:139–151

Kochert G, Branch WD, Simpson CE (1991) RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. Theor Appl Genet 81:565–570

Kochert G, Stalker HT, Gimenes MA et al (1996) RFLP and cytogenetic evidence on the origen and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). Am J Bot 83:1282–1291

Krapovickas A, Gregory WC (1994) Taxonomía del género *Arachis* (Leguminosae). Bonplandia 8:1–186

Krapovickas A, Vanni RO, Pietrarelli JR et al (2009) Las razas de maní de Bolivia. Bonplandia 18:95–189

Krapovickas A, Vanni RO, Pietrarelli JR et al (2013) Las razas de maní de Perú. Bonplandia 22:19–90

Lanfredi M, Congiu L, Garrido-Ramos MA et al (2001) Chromosomal location and evolution of a satellite DNA familyin seven sturgeon species. Chromosome Res 9:47–52

Lavia GI (1996) Estudios cromosómicos en *Arachis* (Leguminosae). Bonplandia 9:111–120

Lavia GI (1998) Karyotypes of *Arachis palustris* and *A. praecox* (Section Arachis), Two Species with Basic Chromosome Number x = 9. Cytologia (Tokyo) 63:177–181

Lavia GI (2000) Chromosomes studies of wild *Arachis* (Leguminosae). Caryologia 53:177–181

Lavia GI, Fernández A (2004) Karyotypic studies in *Arachis hypogaea* L. varieties. Caryologia 57:353–359

Lavia GI, Fernández A (2008) Genome size in wild and cultivated peanut germplasm. Plant Syst Evol 272:1–10

Lavia GI, Ortiz AM, Robledo G et al (2011) Origin of triploid *Arachis pintoi* (Leguminosae) by autopolyploidy evidenced by FISH and meiotic behavior. Ann Bot 108:103–111

Leal-Bertioli SCM, Santos SP, Dantas KM et al (2015) *Arachis batizocoi*: a study of its relationship to cultivated peanut (*A. hypogaea*) and its potential for introgression of wild genes into the peanut crop using induced allotetraploids. Ann Bot 115:237–249

Leitch IJ, Bennett MD (2004) Genome downsizing in polyploidy plants. Biol J Linn Soc 82:651–663

Liu ZL, Zhang D, Hong DY et al (2003) Chromosomal localization of 5S and 18S-5.8S-25S ribosomal DNA

sites in five Asian pines using Fluorescence in situ hybridization. Theor Appl Genet 106:198–204
Loureiro J, Rodriguez E, Costa A et al (2007) Nuclear DNA content estimations in wild olive (*Olea europaea* L. ssp. *europaea* var. *sylvestris* Brot.) and Portuguese cultivars of *O. europaea* using flow cytometry. Genet Resour Crop Evol 54:21–25
Macas J, Požárková D, Navrátilová A et al (2000) Two new families of tandem repeats isolated from genus Vicia using genomic self-priming PCR. Mol Gen Genet 263:741–751
Milla R, Isleib TG, Stalker HT (2005) Taxonomic relationships among Arachis sect. Arachis species as revealed by AFLP markers. Genome 48:1–11
Moretzhon MC, Gouvea EG, Inglis PW et al (2013) A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. Ann Bot 111:113–126
Moretzsohn MDC, Hopkins MS, Mitchell SE et al (2004) Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. BMC Plant Biol. doi:10.1186/14712229411
Mukai Y, Friebe B, Gill BS (1992) Comparison of C-banding patterns and in situ hybridization sites using highly repetitive and total genomic rye DNA probes of 'Imperial' rye chromosomes added to 'Chinese Spring' wheat. Jpn J Genet 67:71–83
Navajas-Pérez R, Del Bosque MQ, Garrido-Ramos MA (2009) Effect of location, organization, and repeat-copy number in satellite-DNA evolution. Mol Gen Genomics 282:395–406
Nielen S, Campos-Fonseca F, Leal-Bertioli SCM et al (2010) FIDEL-a retrovirus-like retrotransposon and its distinct evolutionary histories in the A- and B-genome components of cultivated peanut. Chromosome Res 18:227–246
Nielen S, Vidigal BS, Leal-Bertioli SCM et al (2012) Matita, a new retroelement from peanut: characterization and evolutionary context in the light of the *Arachis* A-B genome divergence. Mol Genet Genomics 287:21–38
Nóbile PM, Gimenes MA, Valls JFM et al (2004) Genetic variation within and among species of genus *Arachis,* section *Rhizomatosae*. Genet Resour Crop Evol 51:299–307
Noma K, Ohtsubo E, Ohtsubo H (1999) Non-LTR retrotransposons (LINEs) as ubiquitous components of plant genomes. Mol Gen Genet 261:71–79
Ortiz AM, Seijo JG, Fernández A et al (2011) Meiotic behavior and pollen viability of tetraploid *Arachis glabrata* and *A. nitida* species (Section Rhizomatosae, Leguminosae): implications concerning their polyploidy nature and seed set production. Plant Syst Evol 292:73–83
Ortiz AM, Seijo G, Lavia GI (2014) *Arachis.* In: Marhold, K (ed), IAPT/IOPB chromosome data 17. Taxon 63:1153
Ortiz AM, Silvestri MC, Lavia GI (2013) Karyotypic studies in wild species of *Arachis* (Leguminosae) belonging to sections Erectoides, Procumbentes and Rhizomatosae. Bol Soc Argent Bot 48:295–300
Ozias-Akins P, Jarret RL (1994) Nuclear DNA content and ploidy levels in the genus *Ipomoea*. J Am Soc Hortic Sci 119:110–115
Paik-Ro OG, Smith RL, Knauft DA (1992) Restriction fragment length polymorphism evaluation of six peanut species within the Arachis section. Theor Appl Genet 84:201–208
Peñaloza APS, Pozzobon MT, Valls JFM (1996) Cytogenetic findings in wild species of *Arachis* (Leguminosae). In: Sociedade Brasileira de Genética (ed) 42 Congresso Nacional de Genetica, Caxambu, September, 1996, Revista Brasilera de Genetica 19:129
Peñaloza APS, Valls JFM (2005) Chromosome number and satellited chromosome morphology of eleven species of *Arachis* (leguminosae). Bonplandia 14:1–8
Pich U, Fritsch R, Schubert I (1996) Closely related *Allium* species (Alliaceae) share a very similar satellite sequence. Plant Syst Evol 202:255–264
Plohl M, Luchetti A, Meštrović N et al (2008) Satellite DNAs between selfishness and functionality: Structure, genomics and evolution of tandem repeats in centromeric heterochromatin. Gene 409:72–82
Pucciariello O, Ortiz AM, Fernández A et al (2013) Análisis cromosómico del híbrido *Arachis pintoi* x *A. repens* (Leguminosae) mediante citogenética clásica y molecular. Bol Soc Argent Bot 48:111–119
Raina SN, Mukai Y (1999) Detection of a variable number of 18S–5. 8S–26S and 5S ribosomal DNA loci by fluorescent in situ hybridization in diploid and tetraploid Arachis species. Genome 59:52–59
Raina SN, Rani V, Kojima T et al (2001) RAPD and ISSR fingerprints as useful genetic markers for analysis of genetic diversity, varietal identification, and phylogenetic relationships in peanut (*Arachis hypogaea*) cultivars and wild species. Genome 44:763–772
Rayburn AL, Biradar DP, Bullock DG et al (1997) Nuclear DNA content diversity in Chinese soybean introductions. Ann Bot 80:321–325
Resslar PM, Stucky JM, Miksche JP (1981) Cytophometric determination of the amount of DNA in *Arachis* (Leguminosae). Amer J Bot 68:149–153
Robledo G, Lavia GI, Seijo G (2009) Species relations among wild *Arachis* species with the A genome as revealed by FISH mapping of rDNA loci and heterochromatin detection. Theor Appl Genet 118:1295–1307
Robledo G, Lavia GI, Seijo G (2010) Genome re-assignment of *Arachis trinitensis* (Sect. Arachis, Leguminosae) and its implications for the genetic origin of cultivated peanut. Genet Mol Biol 33:714–718
Robledo G, Seijo G (2008) Characterization of the *Arachis* (Leguminosae) D genome using fluorescence in situ hybridization (FISH) chromosome markers and total genome DNA hybridization. Genet Mol Biol 31:717–724
Robledo G, Seijo G (2010) Species relationships among the wild B genome of *Arachis* species (section Arachis) based on FISH mapping of rDNA loci and

heterochromatin detection: a new proposal for genome arrangement. Theor Appl Genet 121:1033–1046
Samoluk SS, Chalup L, Robledo G et al (2015a) Genome sizes in diploid and allopolyploid *Arachis* L. species (section Arachis). Genet Resour Crop Evol 62:747–763
Samoluk SS, Robledo G, Podio M et al (2015b) First insight into divergence, representation and chromosome distribution of reverse transcriptase fragments from L1 retrotransposons in peanut and wild relative species. Genetica 143:113–125
Santana SH, Valls JFM (2015) *Arachis veigae* (Fabaceae), the most dispersed wild species of the genus, and yet taxonomically overloocked. Bonplandia 24:139–150
Schmidt T, Heslop- Harrison JS (1998) Genomes, genes and junk: the large-scale organization of plant chromosomes. Trends Plant Sci 3:195–199
Seijo G, Lavia GI, Fernández A et al (2004) Physical mapping of the 5S and 18S–25S rRNA genes by FISH as evidence that *Arachis duranensis* and *A. ipaensis* are the wild diploid progenitors of *A. hypogaea* (Leguminosae). Am J Bot 91:1294–1303
Seijo G, Lavia GI, Fernández A et al (2007) Genomic relationships between the cultivated peanut (*Arachis hypogaea*, Leguminosae) and its close relatives revealed by double GISH. Am J Bot 94:1963–1971
Shirasawa K, Koilkonda P, Aoki K et al (2012) *In silico* polymorphism analysis for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. BMC Plant Biol. doi:10.1186/147122291280
Silva SC, Martins MIG, Santos RC et al (2010) Karyological features and banding patterns in *Arachis* species belonging to the Heteranthae section. Plant Syst Evol 285:201–207
Silvestri MC, Ortiz AM, Lavia GI (2015) rDNA loci and heterochromatin positions support a distinct genome type for "x = 9 species" of section Arachis (*Arachis*, Leguminosae). Plant Syst Evol 301:555–562
Singh AK (1986) Utilization of wild relatives in the genetic improvement of *Arachis hypogaea* L. 8. Synthetic amphidiploids and their importance in interspecific breeding. Theor Appl Genet 72:433–439
Singh AK, Gurtu S, Jambunathan R (1993) Phylogenetic relationships in the genus *Arachis* based on seed protein profiles. Euphytica 74:219–225
Singh AK, Raina SN, Singh AK (1996) Variation in chromosomal DNA associated with the evolution of *Arachis* species. Genome 39:890–897
Singh AK, Simpson CE (1994) Biosystematics and genetic resources. In: Smartt J (ed) The groundnut crop, a scientific basis for improvement. Chapman and Hall, London, pp 96–138
Singh AK, Smartt J (1998) The genome donors of the groundnut/ peanut (*Arachis hypogaea* L.) revisited. Genet Resour Crop Evol 45:113–118
Slamovits CH, Cook JA, Lessa EP et al (2001) Recurrent amplifications and deletions of satellite DNA accompanied chromosomal diversification in South American Tuco-tucos (Genus *Ctenomys*, Rodentia: Octodontidae): a phylogenetic approach. Mol Biol Evol 18:1708–1719
Smartt J, Gregory WC (1967) Interspecific cross-compatibility between the cultivated peanut *Arachis hypogaea* L. and other members of the genus *Arachis*. Oléagineux 22:455–459
Smartt J, Gregory WC, Gregory MP (1978) The genome of *A. hypogaea*. 1. Cytogenetic studies of putative genome donors. Euphytica 27:665–675
Smartt J, Stalker HT (1982) Speciation and cytogenetics of *Arachis*. In: Pattee HE, Young CT (eds) Peanut science and technology. Peanut Res and Educ Soc, Yoakum, pp 21–49
Stalker HT (1991) A new species in section Arachis of peanuts with a D genome. Amer J Bot 78:630–637
Stalker HT, Dhesi JS, Parry DC et al (1991) Cytological and interfertility relationships of *Arachis* section Arachis. Am J Bot 78:238–246
Stalker HT, Moss JP (1987) Speciation, cytogenetics, and utilization of *Arachis* species. Adv Agron 41:1–40
Stalker HT, Wynne JC (1979) Cytology of interspecific hybrids in section Arachis of peanuts. Peanut Sci 6:110–114
Tallury SP, Hilu KW, Milla R et al (2005) Genomic affinities in *Arachis* section Arachis (Fabaceae): molecular and cytogenetic evidence. Theor Appl Genet 111:1229–1237
Temsch EM, Greilhuber J (2000) Genome size variation in *Arachis hypogaea* and *A. monticola* re-evaluated. Genome 43:449–451
Temsch EM, Greilhuber J (2001) Genome size in *Arachis* duranensis: a critical study. Genome 44:826–830
Urdampilleta JD, Pereira de Souza A, Schneider DRS et al (2009) Molecular and cytogenetic characterization of an AT-rich satellite DNA family in *Urvillea chacoensis* Hunz. (Paullinieae, Sapindaceae). Genetica 136:171–177
Valls JFM, Costa LC, Custodio AR (2013) A novel trifoliolate species of *Arachis* (Fabaceae) and further comments on the taxonomic section Trierectoides. Bonplandia 22:91–97
Valls JFM, Simpson CE (2005) New species of *Arachis* (Leguminosae) from Brazil, Paraguay and Bolivia. Bonplandia 14:35–63
Wendel JF, Cronn RC, Johnston JS et al (2002) Feast and famine in plant genomes. Genetica 115:37–47
Zhang L, Xu C, Yu W (2012) Cloning and Characterization of Chromosomal Markers from a Cot-1 Library of Peanut (*Arachis hypogaea* L.). Cytogenet Genome Res 137:31–41
Zhang L, Yang X, Tian L et al (2016) Identification of peanut (*Arachis hypogaea*) chromosomes using a fluorescence in situ hybridization system reveals multiple hybridization events during tetraploid peanut formation. New Phytol 211:1424–1439

# Germplasm Characterization and Trait Discovery in Peanut

5

Boshou Liao

**Abstract**

Peanut (*Arachis hypogaea* L.) is an important oilseed and cash crop worldwide. The peanut germplasm is fundamental to genetic enhancement for improved cultivars. A lot of germplasm accessions of the cultivated peanut and wild *Arachis* species have been assembled and conserved in many countries, with ICRISAT, USDA, and OCRI-CAAS being the major conserving agencies. Besides a lot of peanut germplasm characterization work conducted by various scientists in the world, remarked progress has been achieved in the past two decades in assessing the genetic diversity within *A. hypogaea* after the peanut core and mini core collections were selected in the USA, ICRISAT, and China. With extensive and intensive germplasm characterization, elite peanut accessions with desirable traits have been identified for further breeding and other research purposes. Research priorities of trait discovery and genetic enhancement have been given to yield-related characters, resistance to late leaf spot, early leaf spot, rust, tomato spotted wilt virus, groundnut rosette virus, bacterial wilt, nematodes and aflatoxin contamination, tolerance to drought, and quality-related characters such as oil content, protein content, and fatty acid components. With the development of genomic tools, molecular approaches have been widely applied in peanut germplasm characterization and trait discovery.

## 5.1 Introduction

Peanut or groundnut (*Arachis hypogaea* L.) is an important oilseed and cash crop worldwide, grown in more than 100 countries (FAO 2015). It

B. Liao (✉)
Oil Crops Research Institute (OCRI) of Chinese Academy of Agricultural Sciences (CAAS), Wuhan 430062, China
e-mail: lboshou@hotmail.com


R.K. Varshney et al. (eds.), *The Peanut Genome*, Compendium of Plant Genomes, DOI 10.1007/978-3-319-63935-2_5

is crucial for assuring food supply, increasing farmers' income, and promoting sustainable agriculture in many developing countries. Currently, China, India, and the United States of America (USA) are among the top peanut producers in the world. In China, peanut has become the largest oilseed crop in terms of production even where its sowing area is less than rapeseed and soybean (FAO 2015). Varietal improvement has been among the key strategies for developing peanut industry in many countries. Peanut germplasm collection, evaluation, and discovery of key traits are fundamental for developing improved varieties.

Modern breeding objectives in peanut, like in most other crops, include high yield, improved quality, multiple resistances to several biotic and abiotic stresses, early maturity, higher efficiency of water and fertilizer utilization, traits suitable to mechanized practices in the whole production chain, and reduced food safety risks. For all these objectives, breeding programs could only be effectively carried out unless desirable germplasm lines with special traits are identified and available as parents. As in many other crops, the peanut germplasm includes naturally evolved landraces, improved cultivars developed through artificial breeding and selection, breeding lines with special traits, introduced exotic genetic stock, wild relatives, induced mutants through physical or chemical agents, and genetic stocks produced through other approaches. Effective collection, conservation, evaluation, documentation, utilization, and other related basic biological studies have been among the priorities of peanut germplasm research. To date, remarked progress has been achieved worldwide in peanut germplasm collection, characterization, and trait discovery.

## 5.2 Collection and Conservation of Peanut Germplasm

All the species of *Arachis* genus originated in South America, and there are more than 80 individual species within the genus (Barkley et al. 2016). The *Arachis* species including the cultivated peanut, *A. hypogaea* L., can produce geocarpic fruits in soil, which is a unique nature of this special genus. The *Arachis* species are classified as a legume in the plant family Fabaceae (Krapovickas 1969). The cultivated peanut is a tetraploid (2n = 4x = 40), while most wild species in the *Arachis* genus are diploid (2n = 2x = 20) even there are also several tetraploid wild species (Stalker and Simpson 1995). The tetraploid peanut with the genome AABB is believed to have originated from a natural hybrid between two diploid wild species. Recent molecular studies have elucidated that *A. duranensis* and *A. ipaensis* are the wild parents which might have contributed the genome A and B, respectively, to the cultivated peanut (Kochert et al. 1991, 1996; Moretzsohn et al. 2012; Seijo et al. 2007; Stalker 1997). The cultivated peanut is the only species in the genus extensively cultivated and utilized (Stalker and Simpson 1995). The cultivated peanut is classified into two subspecies, *subsp hypogaea* and *subsp fastigiata*, based on the presence or absence of floral axes on the main stem. The *subsp hypogaea* is further divided into two botanic varieties, *var hirsuta* and *var hypogaea*, while the *subsp fastigiata* is further divided into four botanic varieties, *var aequatoriana*, *var fasitigiata*, *var peruviana*, and *var vulgaris*, based on a range of morphological characteristics (Krapovickas and Gregory 1994). However, there are so many intermediates among these botanical types; therefore, the taxonomy of the cultivated species is not always clear (Stalker and Simpson 1995). Extensive collection, characterization, and utilization have also been conducted for wild *Arachis* species. To date, 80 wild *Arachis* species have been described, belonging to nine distinct taxonomic sections based on cross-compatibility, morphological characters, and geographic origin (Krapovickas and Gregory 1994; Valls and Simpson 2005).

There has been a long history since peanut was disseminated from South America to other continents. The discovery of South America by Columbus in 1492 was thought as a key event for transmission of peanuts to other continents, but peanuts might have been shipped out of South America before Europeans reached there (Sun

1998). In South America, the earliest archeological record of peanut comes from Peru and could date back to 3900–3750 years ago (Hammons 1994). However, systematic collection and conversation of peanut germplasm was just initiated since the early twentieth century. Obviously, before the modern agricultural production and research systems were initiated, peanut variety choice, seed maintaining, and dissemination were decided by individual favor of the growers for local or family consumption. Artificial and natural selection have had a great impact on survival of the naturally evolved peanut germplasm. On the other hand, some special varieties might have also survived due to biological or geographic isolation. With the development and extension of modern peanut cultivars, collection and conservation of traditional landraces are crucial for long-term utilization of the naturally evolved germplasm resources. The largest collections of *Arachis* germplasm are housed in the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) in India, the United States, China, and Brazil, although smaller collections do exist in many countries around the world (Barkley et al. 2016).

Throughout the world, the USA first initiated systematic peanut germplasm collection. Currently, the USDA Agricultural Research Service (ARS) Plant Genetic Resources Conservation Unit located in Griffin, Georgia (USA) maintains a large *Arachis* germplasm collection, consisting of 9976 accessions of which 9321 are cultivated peanut accessions and 655 are wild *Arachis* accessions belonging to 73 species (Barkley et al. 2016). This collection is derived from various peanut lines collected from 102 countries. Approximately, 44% of the collection was collected from South America where peanut originated. The cultivated peanut collection and some of the wild species are maintained as seeds, while the wild species those do not produce seed are maintained as colonial plants (Barkley et al. 2016). Besides stored in Griffin, the majority (>95%) of the USDA peanut collection is also backed up at the National Center for Genetic Resources Preservation in Fort Collins, Colorado, USA.

Since the late 1970s, ICRISAT located in Hyderabad, India, has become the international center for global peanut germplasm conservation (Upadhyaya et al. 2001a). To date, ICRSAT has assembled and maintained the world's largest peanut germplasm collection consisting of 15,446 accessions from 92 countries (Barkley et al. 2016). The current ICRISAT collection includes 14,968 accessions of cultivated peanut and 478 accessions of 48 wild *Arachis* species (Barkley et al. 2016). Among the cultivated peanut accessions maintained at ICRISAT, 7172 are traditional cultivars/landraces, 979 are advanced or improved cultivars, 4986 are breeding lines/research materials, and 1831 are genetic stocks derived from various means or resources (Upadhyaya et al. 2014). Besides stored at ICRISAT genebank in Hyderabad, more than 13,900 accessions of the ICRISAT peanut germplasm collection are backed up at the Global Seed Vault at Svalbard, Norway.

There has been a long history of peanut cultivation in China where the natural conditions are much more diversified than any other peanut producing country, which has resulted in diverse peanut germplasm. In some remote regions in China, special peanut landraces could still be seen in local farmer's field. In China, systematic collection and preservation of peanut germplasm was first initiated in the early 1950s (Sun 1998). By 1959, a total of 1239 accessions of cultivated peanut were collected from various provinces, and then, the number of collected accessions increased to 2378 in 1963 (Sun 1998). The collection of peanut landraces in the 1950s and early 1960s also facilitated the research on the classification of cultivated peanut in China (Sun 1998). Compared to the peanut germplasm collection in other countries, more Dragon type (equal to *var hirsuta*) genotypes were collected in China. By 2014, a total of 8439 accessions of cultivated peanut including 4638 landraces from 22 provinces were collected and assembled. The peanut collection is maintained at the Oil Crops Research Institute of Chinese Academy of

Agricultural Sciences (OCRI-CAAS) located in Wuhan for midterm conservation. Meanwhile, most of the collection is also backed up at the National Crop Germplasm Genebank in Beijing. In addition to the cultivated peanut germplasm, 246 wild *Arachis* accessions belonging to 26 species have been introduced from the USA and ICRISAT and characterized for various traits since 1979. A National *Arachis* Nursery was established at OCRI-CAAS in Wuhan for preserving these wild accessions.

## 5.3 Characterization and Evaluation of Peanut Germplasm

As in many other crops, effective characterization and evaluation of peanut germplasm are the prerequisite for efficient utilization in breeding and other research purposes. Reasonable grouping and cataloging based on essential characterization are highly necessary so that breeders and other researchers can focus on those desirable lines with special traits. Worldwide standard for peanut germplasm description, techniques for characterization, and scope of different traits and germplasm information access have been improved with the increase of collected germplasm accessions. In 1992, the International Board for Plant Genetic Resources (IBPGR, now renamed as International Plant Genetic Resource Institute, IPGRI) and ICRISAT jointly developed the Descriptors for Groundnut (IBPGR and ICRISAT 1992; Pittman 1995) that has been extensively applied in different institutions working on peanut in the world. The comparativeness of germplasm characterization among different institutions is much improved by applying the standard descriptors. A similar technical protocol named as Descriptors and Data Standard for Peanut (*Arachis* spp.) was compiled and published in 2006 in China with some modifications of Descriptors for Groundnut (Jiang and Duan 2006). In the above technical protocols, more than 160 descriptors are involved, among which the botanical, morphological, agronomic traits, growth period, resistance to biotic and abiotic stresses, and chemical compositions are of common importance in peanut germplasm characterization.

With an extensive collection of peanut germplasm, characterization and evaluation have been extensively conducted in many institutions worldwide. However, most of the characterization has been carried out based on various natural conditions, experimental facilities, and research aims. For example, 14,952 accessions of cultivated peanut and 292 wild *Arachis* accessions collected in ICRISAT have been basically characterized. More than 9000 accessions have been basically characterized in the USDA collection and more than 8000 accessions characterized in the Chinese collection (Barkley et al. 2016). However, in terms of the huge number of peanut germplasm accessions, the intensive characterization and evaluation in the available collections are far from satisfaction. In most cases, it is very difficult to systematically characterize the entire collection consisting of thousands of accessions for all important characters, which would be one of the reasons why only a very small portion of the collected germplasm lines has been used in breeding (Barkley et al. 2016). To overcome this problem, selecting and characterizing core collection have been a practical approach in most crops including peanut.

Core collection is a concept proposed by Harlan (1972) and Frankel (1984) for developing and using subsets of germplasm collections. In general, a core collection could minimize repetitiveness within the collection and should, to the extent possible, represent the genetic diversity in the crop species (Barkley et al. 2016). Selection of core collections could facilitate easier evaluation of genetic diversity and access to genetic resources with special traits, enhance their utilization in crop improvement, and also simplify genebank management.

Holbrook et al. (1993) developed the first peanut core collection in the world. The available entire peanut accessions collected and conserved in the USA were first stratified by country of origin, and then divided into nine groups according to other additional information available and to the number of accessions per country

of origin. Finally, a core collection consisting of 831 accessions (accounting for 11.18% of the entire collection) was selected. Examination of means and ranges for six morphological variables indicated that this core collection was a nice representative sample of the entire collection, and that the genetic variation expressed for these traits in the entire collection was well preserved in this core collection (Holbrook et al. 1993). After this core collection was developed, several scientists have conducted intensive characterization for various traits, in particular for tomato spotted wilt virus (TSWV) (Anderson et al. 1996), Sclerotinia blight and pepper spot (Damicone et al. 2010), early leaf spot and Cylindrocladium black rot (Isleib et al. 1995), late leaf spot (Holbrook and Anderson 1995), peanut root-knot nematode (*Meloidogyne arenaria*) (Holbrook et al. 2000a, b), Rhizoctonia limb rot (Franke et al. 1999), nutrient-related traits (Dean et al. 2009), yield, and aflatoxin contamination under drought stress (Holbrook et al. 2009). The accessions in the US core collection have also been used to evaluate genetic variation in fatty acid composition (Hammond et al. 1997). Obviously, the development of core collection has greatly contributed to the intensive characterization of the peanut germplasm.

In the USA, a core collection was also developed for the Valencia market type (*A. hypogaea subsp fastigiata var fastigiata*) germplasm consisting of 630 accessions in the USDA peanut collection. These concerned accessions were characterized for 26 descriptors, and certain information was obtained from two seasons' observation (Dwivedi et al. 2008). The accessions were stratified by country of origin, and data on morphological and agronomic descriptors were used for clustering following Ward's method. About 10%, or a minimum of one accession from each cluster and region, were selected to develop a core subset of 77 accessions (accounting for 12.22% of the entire collection). This core collection would be very useful in varietal improvement for the Valencia market type.

From the peanut germplasm collection consisting of 14,310 accessions assembled at ICRISAT, Upadhyaya et al. (2003) selected a core collection consisting of 1704 accessions (accounting for 11.9% of the entire collection). In developing this core collection, the ICRISAT entire peanut collection was first stratified by botanical variety within each subspecies, and then stratified by country of origin. Accessions of the same botanical variety from small and adjacent countries with similar agro-climates were grouped together, and the accessions were divided into 75 groups. The accessions within each group were then clustered using multivariate statistical analysis. Approximately, 10% of the accessions from each cluster were randomly sampled. Among the 1704 accessions in the selected core collection, 910 *subsp fastigiata* and 794 *subsp hypogaea* lines were included (Upadhyaya et al. 2003). After this, the core collection was evaluated for 16 morphological descriptors, oil and protein content in one season, and for 15 agronomic traits in two seasons (Upadhyaya 2003). The phenotypic diversity in the core collection was estimated, and the importance of different descriptor traits in explaining the genetic variation was determined. The results revealed significant variation for morphological and agronomic traits in the peanut core collection. For all the traits involved except for trichomes on leaflet surface and for oil content, the *subsp fastigiata* and *subsp hypogaea* groups differed significantly. The *subsp hypogaea* group possessed significantly greater mean pod length, pod width, seed length, seed width, yield per plant, and 100-seed weight than the *subsp fastigiata* group, whereas it was opposite for plant height, leaflet length, leaflet width, and shelling percentage (Upadhyaya 2003).

Upadhyaya et al. (2001b) also developed a peanut core collection by using 4738 germplasm accessions collected from 21 Asian countries including 267 *var fastigiata*, 2414 *var vulgaris*, and 2057 *var hypogaea* lines. In developing this Asian core collection, the 4738 accessions were stratified by country of origin within each of three botanical varieties. Data on 15 morphological descriptor traits including growth habit, branching pattern, stem (color, trichomes), leaf (color, shape, trichomes, flower, streak, and peg

(color)), pod (beak, constriction, and reticulation), and seed (per pod, color) were used for clustering. Ten percent from each cluster or a minimum of one accession per cluster was randomly selected. The final Asian core collection consisted of 504 accessions (accounting for 10.64% of the entire collection), including 274 accessions of *subsp fastigiata* (29 accessions of *var fastigiata*, 245 accessions of *var vulgaris*), and 230 of *subsp hypogaea var hypogaea*. The Asian core collection, along with four control cultivars, was evaluated in multi-environments for 22 agronomic traits to select diverse superior germplasm accessions for use as parents in breeding programs. About 60 lines were selected based on the evaluation and could be used in the peanut improvement programs to broaden the genetic base of cultivars (Upadhyaya et al. 2001b).

To represent the entire peanut germplasm collection assembled in China, a core collection was selected by using the available information (Jiang et al. 2008b). The core collection was selected from the entire collection consisting of 6390 accessions which were assembled and basically described up to 2005. The entire collection was first divided into five groups by botanical types and into 32 subgroups by origin, and then further into 258 variety clusters based on the available data derived from various characterization. The number of germplasm accessions within each cluster was 21 for Valencia, 100 for Spanish, 100 for Virginia, 19 for Dragon, and 18 for irregular (improved cultivars) types. In each cluster, 5–10% was sampled randomly. A total of 576 accessions were selected to form the core collection, accounting for 9.01% of the entire collection. Both in the selected core collection and the entire collection, the diversity indexes were higher in Virginia and Dragon types and relatively lower in irregular type. The difference in diversity index between the entire collection and the core collection was not significant in each botanical type, indicating that the selected core collection could well represent the diversity in the entire collection. Evaluation of this core collection indicated that the Chinese peanut collection might be an important source of diversity in *var hirsuta* and *var vulgaris*. With the development of the peanut core collection, intensive evaluation has also been conducted in China for resistance to bacterial wilt, foliar diseases, seed infection by *Aspergillus flavus*, aflatoxin production, and fatty acid composition (Jiang et al. 2014).

Although core collection can represent the diversity in the entire collection by using 10% samples which can greatly facilitate efficiency of evaluation of germplasm and of crop improvement, the number of accessions in core collection is still more than 570 for the Chinese core collection (Jiang et al. 2008b), more than 800 for the USDA core collection (Holbrook et al. 1993; Holbrook and Anderson 1993), and even more than 1700 for the ICRISAT core collection (Upadhyaya et al. 2003). In such size of peanut samples, it is still less feasible for scientists to characterize most traits precisely, in particular for aflatoxin resistance, drought tolerance, and most quality traits which are generally difficult to characterize. Under this background, Upadhyaya et al. (2002) postulated the concept of "mini core", wherein approximately 10% of core collection is subsampled (or 1% of the entire collection) to represent global diversity in the species. Upadhyaya et al. (2002) developed the first peanut mini core collection. This mini core collection, consisting of 184 accessions, was selected from the 1704 accessions of the core collection (Upadhyaya et al. 2003). Before selecting the mini core collection, all the accessions in the core collection were evaluated for the morphological, agronomic, and quality traits under field conditions at ICRISAT located in Patancheru of India. A phenotypic distance matrix was created by calculating differences between each pair of accessions for each of the 47 traits involved. This distance matrix was then subjected to hierarchical cluster analysis, and based on the analysis, the accessions in the core collection were grouped into 77 clusters. From each cluster, approximately 10% of the accessions were randomly selected to form the mini core subset, but at least one accession was included from each cluster even if the accession number was 10 or less. The mini core collection

(184 accessions) accounted for 10.8% of the core collection (Upadhyaya et al. 2002).

In the USA, shortly after the peanut core collection was developed in 1993 (Holbrook et al. 1993), selection of a smaller subset germplasm was thought as highly necessary for germplasm evaluation on certain traits that were difficult or expensive to characterize. Based on the characterization data of the core collection, Holbrook and Dong (2005) used multivariate analysis to select a mini core collection, to represent the US peanut diversity. The mini core collection consisted of 112 accessions, accounting for 13.5% of the core collection. The mini core collection was believed to be useful in improving the efficiency of identifying donors for desirable traits in the core collection (Holbrook and Dong 2005; Chenault Chamberlin et al. 2010), and the mini core collection has been evaluated for several new traits (Wang et al. 2011, 2013). More recently, the USDA mini core collection was further purified (Chen et al. 2013).

A peanut mini core collection consisting of 298 accessions was selected in China based on characterizing the major traits of the selected core collection (Jiang et al. 2008a, b). Genetic diversity in the peanut mini core collections of China and ICRISAT was compared by using SSR markers, and it was concluded that considerable genetic difference existed between the Chinese and some ICRISAT peanut accessions especially with the *var aequatoriana* genotype, ICG 12625. The genetic diversity was greater among the Chinese peanut mini core than that among the ICRISAT mini core collection in terms of similarity coefficient and genetic diversity index (Jiang et al. 2010). However, the number of accessions in the Chinese mini core collection was still too large. Therefore, for more convenient and efficient use, a mini-mini core consisting of 99 accessions was selected based on information of 21 morphological traits characterized in the core collection (Jiang et al. 2013). It was demonstrated that there were no significant differences between the core and the mini-mini core collections in 20 out of the 21 morphological traits involved. Furthermore, the mini-mini core collection captured the ranges of all of the 21 traits displayed in the core collection. The newly developed mini-mini core collection was assessed for resistance to bacterial wilt caused by *Ralstonia solanacearum*. Two accessions showing a high level of resistance to bacterial wilt were identified, demonstrating the usefulness of the mini-mini core collection.

Characterization of wild *Arachis* species is generally crucial for peanut varietal improvement. To date, a lot of synthesized tetraploid peanut was developed at ICRISAT by combining putative A and B genome *Arachis* species. Many of the interesting traits in the diploid species still exist in the synthesized tetraploid genetic stocks. The newly synthesized tetraploid materials possessing certain interesting traits would be hopeful in transferring the useful traits into high-yielding peanut cultivars (Mallikarjuna et al. 2012).

Besides phenotypical characterization of morphological, agronomic and quality traits of peanut germplasm, some molecular approaches have been extensively used in analyzing genetic diversity, germplasm characterization, trait mapping, and molecular breeding (Upadhyaya et al. 2008). Molecular markers including AFLPs (amplified fragment length polymorphism), RAPDs (random amplified polymorphic DNA), RFLPs (restriction fragment length polymorphism), SNPs (single nucleotide polymorphisms), and SSRs (simple sequence repeats) have been fairly used by peanut researchers (Kottapalli et al. 2007; Pandey et al. 2012). Assessing the level of genetic diversity in peanut germplasm accessions through fingerprinting with suitable molecular markers can provide necessary information on relatedness among individuals and genetic redundancy or diversity, and can reveal misidentified materials, population structure, and hybrid origins (Barkley et al. 2007; Wang et al. 2011). Molecular markers have been extensively utilized to assess inter- and intraspecific genetic variation in cultivated and wild peanut germplasm (Pandey et al. 2012; Huang et al. 2012). Based on molecular characterization, many researchers have found the diversity levels to be quite low within the cultivated peanut (Liang et al. 2009; Moretzsohn et al. 2004; Pandey et al. 2012; Stalker and

Mozingo 2001). Further integration of the diverse genomics tools with ongoing conventional breeding techniques is expected to facilitate the development of improved peanut cultivars more efficiently, in particular for certain traits that are difficult to enhance by conventional approaches.

## 5.4 Identification of Key Traits in Peanut

For peanut germplasm research, it is a crucial task to identify special germplasm lines harboring special traits needed in varietal improvement through properly characterizing the assembled germplasm accessions. In view of efficient breeding utilization, comprehensive evaluation of germplasm accessions possessing similar target traits for identifying genotypes with the better overall genetic background is also important. Among the peanut producing countries, the natural, economic, and agricultural conditions for peanut production are quite variable while the utilization of peanuts may be also much diverse; therefore, the objectives of varietal improvement in different countries or regions are quite different. However, high yielding, early maturity, disease resistance, stress tolerance, and improved quality are among the common interesting or target traits across locations and institutions.

As in other crops, yield in peanut is a complex trait related with several components. Most of the yield-related traits are quantitatively inherited with different heritability. In general, large pod or seed is a key component for high yield in peanut. The number of reproductive branches and the number of pods per plant are also crucial for high yield, even these two components are highly affected by environmental factors. In terms of pod or seed size and number of pod per plant, accessions of *subsp hypogaea* are much more diverse than the *subsp fastigiata*. In China, the records of high yield cultivation of peanut with a yield over 8 ton/ha were only achieved by using large-podded Virginia type varieties (Yu 2011). Shelling percentage in peanut, a trait related to seed yield, is variable among the germplasm collection and relatively stable equal to a qualitative trait. Most of the yield-related traits have been characterized in most countries. In the Chinese peanut core collection, the largest average pod weight (per 100 pods) was in the Intermediate type (derived from hybrid between botanic types) as 173.8 g followed by Virginia type (161.3 g), Dragon type (142.3 g), Spanish type (133.5 g), and Valencia type (124.0 g). In the history of peanut cultivar improvement, larger pods or seeds have been selected for high yield (Yu 2011). In the core collection, the shelling percentage ranged from 37.0 to 85.9% with an average as 71.3%. Among various botanical types, the highest shelling percentage was in the Dragon type (73.0%) followed by Spanish (72.3%), Intermediate (70.7%), Virginia (70.4%), and Valencia (70.4%) types (Barkley et al. 2016).

Early maturity is an important trait in peanut. Cultivars with early maturity and optimum yield can not only meet with a special cropping system for rotation, but also an important feature for avoiding environmental stress such as end-season drought. The ICRISAT peanut core collection was evaluated with the aim to identify the diverse early maturity accessions, which resulted in the identification of 21 early-maturing lines (Upadhyaya et al. 2006). These early-maturing landraces can produce 12.6% more yield at 75 days after sowing (DAS) and 8.4% at 90 DAS than the mean of three early-maturing control cultivars (Chico, Gangapuri, and JL 24). Four early-maturing landraces (ICG 4558 (India), ICG 4890 (Argentina), ICG 9930 (Zimbabwe), and ICG 11605 (Bolivia)) with predominantly three to four seeds per pod were newly identified as additional sources for breeding confectionery peanut cultivars. These diverse early-maturing accessions from different locations are agronomically superior, and therefore are likely to provide better opportunities in developing early maturity cultivars suitable for different geographic regions.

Resistance to diseases is an important breeding objective for peanut. Discovery of disease-resistant germplasm has been a priority in many peanut germplasm and breeding

programs in the world. Late leaf spot (LLS) caused by *Phaeoisariopsis personata* has been the most widespread foliar disease of peanut in the world (Singh et al. 1997). This disease could be almost seen in all the peanut producing regions in the world every year. Screening and evaluation of peanut germplasm accessions for resistance to late leaf spot have attracted much research efforts in many countries in particular in ICRISAT, India, China, USA, and several African nations. Early leaf spot (ELS) caused by *Cercospora arachidicola* is another widespread foliar disease affecting peanut in many regions, but the epidemic area of early leaf spot is much smaller than the late leaf spot. Therefore, the large-scale screening for resistance to ELS has been only conducted in several locations where the natural occurrence of the disease is serious. For peanut, another foliar disease occurring in most tropic and subtropic regions has been rust caused by *Puccinia arachidis*. Even the severity of peanut rust in many regions has declined since the late 1960s, the disease is still serious in many tropic producing regions such as India, tropic Africa, and Southeast Asia. The large-scale screening for rust resistance in peanut germplasm has been carried out in ICRISAT, the USA, and China (Singh et al. 1997, 2003). Since peanut became one of the mandate crops in ICRISAT, screening resistance for late leaf spot and rust were continuously conducted for many years in Patancheru, Hyderabad where the natural occurrence of these two foliar diseases was stably serious. More than 10,000 accessions have been systematically evaluated for their reaction to these two diseases. Special techniques for resistance screening including growing "infector rows" 1 week earlier of the germplasm lines to be tested, additional inoculation by spraying pathogen suspension in the field, and imposing necessary field irrigation for assuring moisture for disease development were empirically established. In addition to field screening, techniques for whole plant evaluation in the greenhouse and in vitro screening of detached leaves were also well developed. A 1–9 disease score was standardized for grading the disease in both field and greenhouse (Subrahmanyam et al. 1995). In addition to disease score in field or greenhouse, components of resistance including latent period, number of lesions, size of lesions, and destroyed leaf area were used to evaluate and distinguish the relative difference in resistance among different germplasm resources. The work conducted at ICRISAT has greatly contributed to the genetic improvement of foliar disease resistance in many countries. The identified germplasm lines with high-level resistance to late leaf spot, early leaf spot, and rust have been reviewed by Singh et al. (1997) and Barkley et al. (2016). Based on the available evaluation of resistances to the above mentioned three key foliar diseases in peanut germplasm, it could be concluded that most genotypes harboring high level of resistance in the cultigen are *var fastigiata* accessions collected from South America. Many accessions of wild *Arachis* species are highly resistant to leaf spot and even immune to rust. Through interspecific hybridization, high-level resistance to foliar diseases has been successfully transferred into the cultigens (Mallikarjuna 2002). However, the *var fastigiata* resistant genotypes identified have been less successfully utilized in breeding because of genetic linkage between the resistance and some undesirable agronomic traits including deep reticulation of pod, thick shell, poor seed color, and low shelling percentage (Liao 2014). Fortunately, some interspecific hybrid derivatives have been successfully used in breeding with high level of resistance to foliar diseases and better agronomic traits (Liao 2014; Mallikarjuna 2002; Mallikarjuna et al. 2012; Singh et al. 1997).

Bacterial wilt (BW) caused by *Ralstonia solanacearum* has been the only bacterial disease in peanut with economic importance in the world (Jianwei et al. 2010; Liao 2014). The disease is a big constraint to peanut production in several countries including China, Indonesia, Vietnam, and Uganda. Unlike most other diseases of peanut, bacterial wilt could only be effectively controlled by planting resistant cultivars as no efficient chemical approach is feasible and available. In Indonesia, breeding of bacterial wilt-resistant peanut cultivars dated back to early 1920s. In China, more than 6300 accessions of

peanut germplasm have been tested for resistance, and totally, more than 140 genotypes with high-level resistance were identified in five botanic types with more resistant lines (accounting for 60%) from the *var hirsuta* type (Liao 2014). High level of resistance to bacterial wilt has also been identified in many wild *Arachis* relative accessions (Stalker et al. 2013), and the resistance has been successfully transferred into cultivated peanut (Liao 2014).

Tomato spotted wilt virus (TSWV) is an important disease in peanut production in the USA (Culbreath et al. 1992). A special resistant accession, PI 203396, was identified from the peanut germplasm collected from Brazil (Barkley et al. 2016). Extensive utilization of this genotype in breeding has resulted in improved cultivars in the USA and contributed to decreased yield loss in production (Isleib et al. 2001). In the USA, Southern Runner was the first commercial peanut cultivar with a moderate level of resistance to TSWV. Like Southern Runner, other TSWV-resistant cultivars such as Georgia Green, C-99R, and Florida MDR98 all have PI 203396 as their direct or indirect parent and were also relatively resistant to late leaf spot. Tifguard has been developed with resistance to both root-knot nematode and TSWV by hybridizing a TSWV-resistant cultivar with a nematode-resistant cultivar (Holbrook et al. 2008).

Groundnut rosette virus (GRV) is a serious peanut disease in several African countries. With efforts made by ICRISAT scientists, resistance to GRV has been identified from some interspecific hybrid derivatives and incorporated into high-yielding peanut cultivars (Subrahmanyam et al. 2001). Several GRV-resistant cultivars have been developed by using the resistant interspecific derivatives. This progress has greatly contributed to peanut production in diseased regions in Africa.

Peanut root-knot nematode is an important constraint to peanut production in several locations in the world. In China, more than 6300 accessions of peanut germplasm were once tested for their reaction to nematode, from which only three lines were identified as resistant (Yu 2011). Among the wild *Arachis* species, high level of resistance or immunity has been identified in *A. cardenasii*, *A. batizocoi*, and *A. diogoi* (Holbrook et al. 2000a). Through interspecific hybridization, the nematode resistance has been transferred into the cultivated peanut and several resistant cultivars with desirable agronomic traits have been developed (Simpson et al. 1993).

Aflatoxin contamination is the key factor affecting peanut industry development and food safety worldwide. Among the peanuts, there are at least two mechanisms to resist aflatoxin contamination: one is resistance to infection of *Aspergillus flavus* or *A. parasiticus* and another is resistance to toxin production after infection in the kernel (Liao 2014; Nigam et al. 2009) even the resistance in peanut is not very high. In addition, variation does exist in peanut shell in resistance to pre- and postharvest fungal infection. Even screening for aflatoxin resistance in peanut has been relatively limited because of difficulty and high cost in evaluation, many resistant genotypes have been reported (Barkley et al. 2016; Holbrook et al. 2009; Liao 2014; Nigam et al. 2009). Nigam et al. (2009) concluded that ICG 1326, ICG 3263, ICG 3700, ICG 4749, ICG 4888, ICG 7633, and ICG 9407 were of special value in aflatoxin resistance breeding program. Only after the core or mini core collections were developed, the diversity of aflatoxin resistance in cultivated peanut was fairly assessed (Jiang et al. 2010) identified 16 peanut accessions with resistance to seed infection or to aflatoxin production from the ICRISAT mini core and the Chinese core collections. From the ICRISAT mini core, seven accessions including ICGs 13603, 1415, 14630, 3584, 5195, 6703, and 6888 were identified as resistant to preharvest aflatoxin contamination over 6 years (2008–2013), and these accessions can be potential sources for investigating mechanisms of resistance and further used in developing resistant cultivars (Waliyar et al. 2016). Wang et al. (2016) assessed the molecular aspects for postharvest peanut seed in response to aflatoxin production and found Zhonghua 6 has active response to resist aflatoxin production.

Around the world, there are several soil-borne fungi diseases causing plant blight or pod rot in peanut. These soil-borne diseases can cause serious yield loss, reduce quality grade, and even increase the risk of mycotoxin contamination. As soil-borne diseases are difficult to control, genetic enhancement for resistant cultivar is crucial. However, screening for resistance to the soil-borne diseases in peanut has been limited due to the complex host–pathogen interactions plus complicated environmental effects. Therefore, identification of resistance to soil-borne fungi diseases should be among the further priorities in peanut genetic enhancement.

Drought stress is a common constraint to peanut production in several regions in the world. Breeding for drought tolerance has been among the priorities of peanut improvement (Upadhyaya 2005). Extensive research efforts have been made in ICRISAT for identifying drought tolerance in peanut germplasm (Hamidou et al. 2012), and identified genotypes have been used in developing cultivars. Efforts have been also made in identifying QTLs related to drought tolerance in peanut (Ravi et al. 2011). Improved drought tolerance of peanut can help in sustaining pod yield and reducing aflatoxin contamination risk (Wright et al. 1994; Reddy et al. 2003). Drought tolerance has been reported to be associated with field aflatoxin contamination of peanut (Arunyanark et al. 2009; Girdthai et al. 2010). However, Hamidou et al. (2014) observed that there was no direct relationship between drought tolerance and aflatoxin contamination even drought stress did increase contamination in the field. Drought tolerance in peanut may be improved through enhancing capability of extracting water from deep soil or enhancing water use efficiency of the plant, or both (Hebbar et al. 1994). Drought tolerance is associated with root weight, length, and size. Selecting large root volume is effective in breeding for drought tolerance (Nigam et al. 2005; Sun 1998). Dwarfed plants with more nodes, thick leaflets, and more hair on leaf surface are also meaningful traits (Sun 1998). Other physiological traits are also associated with drought tolerance including stomatal conductance, surface wax, water potential, membrane stability, accumulation of cytokinins, and evapotranspiration. Normally, selection of drought tolerance in peanut can be based on the performance of morphological and physiological traits. In conventional breeding programs, biomass and pod yield under drought stress have been used for drought tolerance assessment. For physiological selection approaches, some are more expensive than others. Specific leaf area and SPAD Chlorophyll Meter Reading (SCMR) can be inexpensively and easily measured; thus, it has been used in breeding for drought resistance in Australia, ICRISAT, and India (Nageswara Rao et al. 2000). Arunyanark et al. (2008) reported a promising selection index for drought tolerance based on chlorophyll stability. Drought tolerance traits including specific leaf area (SLA) and root length density (RLD) could be contributing to resistance to aflatoxin contamination (Arunyanark et al. 2009). In the USA, C76–16 is identified with improved resistance to drought as well as to aflatoxin contamination (Holbrook et al. 2009). Considerable progress in breeding drought tolerant peanut has been made at ICRISAT, several Indian institutions, and Australia (Nigam et al. 2005).

Quality traits have been among the priorities of peanut varietal improvement in most countries. As a major oilseed crop in developing countries, increasing the oil content of peanut cultivars is crucial for higher oil yield and production efficiency (Liao et al. 2008, 2010). Besides oil content, higher protein content is also a breeding objective in peanut. In addition, the fatty acid composition is very important for the quality of oil and other products and for prolonged shelf life. Based on the characterization data on peanut germplasm accessions including the core and mini core collections, the oil content ranged from 32 to 62%, and the protein content ranged from 12 to 37% (Jiang and Ren 2006; Liao 2014). In China, 55 peanut germplasm accessions have been identified to possess oil content over 57%, and five accessions over 59% (Yu 2011). High oil content has been identified in many wild *Arachis* accessions (Jiang et al. 2010). High oil interspecific

hybrid derivatives have been developed and used in breeding (Liao 2014).

Increasing the content of oleic acid and decreasing linoleic acid and even some saturated fatty acids have been an important objective in peanut breeding. The variation in fatty acids among the cultivated peanut germplasm has been well evaluated and documented (Barkley et al. 2016; Jiang and Ren 2006; Upadhyaya et al. 2012). In general, the germplasm accessions of *subsp hypogaea* possess higher oleic acid. Among about 6000 peanut germplasm accessions tested in China, 22 accessions have been identified to have oleic acid higher than 67%, most belonging to *var hypogaea* and *var hirsuta* in *subsp hypogaea* (Yu 2011). Norden et al. (1987) tested fatty acid composition of 500 peanut genotypes and fortunately obtained two lines with 80% oleic acid and very low linoleic acid (2%). Since the extra high oleic mutant was found, genetic improvement of oleic acid by crossing traditional parents with relatively high oleic acid has become less important. By using molecular selection (Barkley et al. 2010, 2011) and backcrossing approach, the high oleate trait can be easily identified and quickly transferred into the high-yielding genetic background without changing other major traits. In the available high oleate peanut, the palmitic acid is reduced by more than 40%. If the saturated acids in the high oleate peanut could be reduced further, the quality of peanut and peanut products will be more competitive in markets in the future.

## 5.5 Conclusion

Most peanut germplasm materials in the world have been collected and conserved, even further collection from certain countries in South America is needed to rescue special variants both of the cultivated peanut and wild *Arachis* species. By using the core and mini core collection strategies, along with new techniques, standardized characterization and evaluation have been extensively and intensively conducted for the major peanut germplasm collections in the world including ICRISAT, USA, and China. Germplasm characterization and discovery of key traits have contributed greatly to varietal improvement in peanut, which dramatically promoted the production. Furthermore, genomic techniques have been extensively used in germplasm characterization and trait discovery of peanut. With the accomplishment of peanut genome sequencing and annotation of more key genes, trait discovery and genetic improvement in peanut are expected to progress more quickly than ever before.

## References

Anderson WF, Holbrook CC, Culbreath AK (1996) Screening the peanut core collection for resistance to tomato spotted wilt virus. Peanut Sci 23:57–61

Arunyanark A, Jogloy S, Akkasaeng C, Vorasoot N, Kesmala T, Nageswara Rao RC, Wright GC, Patanothai A (2008) Chlorophyll stability is an indicator of drought tolerance in peanut. J Agron Crop Sci 194:113–125

Arunyanark A, Jogloy S, Wongkaew S, Akkasaeng C, Vorasoot N, Wright GC, Nageswara Rao RC, Patanothai A (2009) Association between aflatoxin contamination and drought tolerance traits in peanut. Field Crops Res 114:14–22

Barkley NA, Dean RE, Pittman RN, Wang ML, Holbrook CC, Pederson GA (2007) Genetic diversity of cultivated and wild-type peanuts evaluated with M13-tailed SSR markers and sequencing. Genet Res 89:93–106

Barkley NA, Chamberlin Chenault KD, Wang ML, Pittman RN (2010) Development of real-time PCR genotyping assay to identify high oleic acid peanuts (*Arachis hypogaea* L.). Mol Breed 25:541–548

Barkley NA, Wang ML, Pittman RN (2011) A real-time PCR genotyping assay to detect *FAD2A* SNPs in peanuts (*Arachis hypogaea* L.). Electron J Biotech. doi:10.2225/vol14-issue1-fulltext-12

Barkley NA, Upadhyaya HD, Liao B, Holbrook CC (2016) Global resources of genetic diversity in peanut. In: Stalker HT, Wilson RF (eds) Peanuts: genetics, processing, and utilization. academic press and AOCS press, pp 67–109. ISBN: 9781630670382

Chen CY, Barkley NA, Wang ML, Holbrook CC, Dang PM (2013) Registration of purified accessions for the U.S. peanut mini-core germplasm collection. J Plant Regist 8:77–85

Chenault Chamberlin KD, Melouk HA, Payton ME (2010) Evaluation of the U.S. peanut mini core collection using a molecular marker for resistance to *Sclerotinia minor* jagger. Euphytica 172:109–115

Culbreath AK, Todd JW, Demski JW, Chamberlin JR (1992) Disease progress of spotted wilt in peanut

cultivars florunner and southern runner. Phytopathology 82:766–771
Damicone JP, Holbrook CC, Smith DL, Melouk HA, Chenault KD (2010) Reaction of the core collection of peanut germplasm to sclerotinia blight and pepper spot. Peanut Sci 37:1–11
Dean LL, Hendrix KW, Holbrook CC, Sanders TH (2009) Content of some nutrients in the core of the peanut germplasm collection. Peanut Sci 36:104–120
Dwivedi SL, Puppala N, Upadhyaya HD, Manivannan N, Singh S (2008) Developing a core collection of peanut specific to Valencia market type. Crop Sci 48:625–632
FAO ( 2015) FAO web. http://www.fao.org/
Franke MD, Brenneman TB, Holbrook CC (1999) Identification of resistance to rhizoctonia limb rot in a core collection of peanut germplasm. Plant Dis 83:944–948
Frankel OH (1984) Genetic perspectives of germplasm conservation. In: Arber WK, Llimensee K, Peacock WJ, Starlinger P (eds) Genetic Manipulation: impact on man and society. Cambridge University Press, Cambridge, UK, pp 161–170
Girdthai T, Jogloy S, Vorasoot N, Akkasaeng C, Wongkaew S, Holbrook CC, Patanothai A (2010) Associations between physiological traits for drought tolerance and aflatoxin contamination in peanut genotypes under terminal drought. Plant Breed 129:693–699
Hamidou F, Ratnakumar P, Halilou O, Mponda O, Kapewa T, Monyo E, Faye I, Ntare BR, Nigam SN, Upadhyaya HD, Vadez V (2012) Selection of intermittent drought tolerant lines across years and locations in the reference collection of groundnut (*Arachis hypogaea* L.). Field Crops Res 126:189–199
Hamidou F, Rathore A, Waliyar F, Vadez V (2014) Although drought intensity increases aflatoxin contamination, drought tolerance does not lead to less aflatoxin contamination. Field Crops Res 156:103–110
Hammons RO (1994) The origin and history of the groundnut. In: Smartt J (ed) The groundnut crop: a scientific basis for improvement. Chapman and Hall, London, pp 24–42
Hammond EG, Duvick D, Wang T, Dodo H, Pittman RN (1997) Survey of fatty acid composition of peanut (*Arachis hypogaea*) germplasm and characterization of their epoxy and eicosenoic acids. J Am Chem Soc 74:1235–1239
Harlan JR (1972) Genetic resources in sorghum. In: Rao NGP, House LR (eds) Sorghum in the seventies. Oxford & IBH Publishing Co., New Delhi, pp 1–13
Hebbar KB, Sashidhar VR, Udayakumar M, Devendra R, Nageswara Rao RC (1994) A comparative assessment of water use efficiency in groundnut (*Arachis hypogaea*) grown in containers and in the field under water-limited conditions. J Agri Sci 122:429–434
Holbrook CC, Anderson WF (1993) Minimum descriptor information on a core collection of peanut. Agron Abstr 189
Holbrook CC, Anderson WF, Pittman RN (1993) Selection of a core collection from the U.S. germplasm collection of peanut. Crop Sci 33:859–861
Holbrook CC, Anderson WF (1995) Evaluation of a core collection to identify resistance to late leafspot in peanut. Crop Sci 35:1700–1702
Holbrook CC, Stephenson MG, Johnson AW (2000a) Level and geographical distribution of resistance to *Meloidogyne arenaria* in the U.S. peanut germplasm collection. Crop Sci 40:1168–1171
Holbrook CC, Timper P, Xue HQ (2000b) Evaluation of the core collection approach for identifying resistance to *Meloidogyne arenaria* in peanut. Crop Sci 40:1172–1175
Holbrook CC, Dong W (2005) Development and evaluation of a mini-core collection for the U.S. peanut germplasm collection. Crop Sci 45:1540–1544
Holbrook CC, Timper P, Culbreath AK, Kvein CK (2008) Registration of 'Tifguard' peanut. J Plant Regist 2:92–94
Holbrook CC, Guo BZ, Wilson DM, Timper P (2009) The U.S. breeding program to develop peanut with drought tolerance and reduced aflatoxin contamination. Peanut Sci 36:50–53
Huang L, Jiang H, Ren X, Chen Y, Xiao Y, Zhao X, Tang M, Huang J, Upadhyaya HD, Liao B (2012) Abundant microsatellite diversity and oil content in wild *Arachis* species. PLoS One 7:e50002
IBPGR, ICRISAT (1992) Descriptors for Groundnut. International Board of Plant Genetic Resources. Rome, Italy. ICRAST, Patancheru, AP, India
Isleib TG, Beute MK, Rice PW, Hollowell JE (1995) Screening of the peanut core collection for resistance to cylindrocladium black rot and early leaf spot. Proc Am Peanut Res Educ Soc 27:25
Isleib TG, Holbrook CC, Gorbet DW (2001) Use of plant introductions in peanut cultivar development. Peanut Sci 28:96–113
Jiang H, Ren XP (2006) Genetic diversity of peanut resource on morphological characters and seed chemical components in China. Chin J Oil Crop Sci 28:421–426
Jiang HF, Duan NX (2006) Descriptors and Data Standard for Peanut (*Arachis* spp.). China Agricultural Press, Beijing. ISBN: 7-109-10911-9 (in Chinese)
Jiang HF, Ren XP, Huang JQ, Liao BS, Lei Y (2008a) Establishment of a peanut mini core collection in China and exploration of new resource with high oleate. Chin J Oil Crop Sci 30:294–299
Jiang HF, Ren XP, Liao BS, Huang JQ, Lei Y, Chen BY, Guo BZ, Holbrook CC, Upadhyaya HD (2008b) Peanut core collection established in China and compared with ICRISAT mini core collection. Acta Agron Sin 34:25–30
Jiang HF, Ren XP, Wang SY, Zhang XJ, Huang JQ, Liao BS, Holbrook CC, Upadhyaya HD (2010) Development and evaluation of peanut germplasm with resistance to *Aspergillus flavus* from core collection. Acta Agron Sin 36:428–434

Jiang HF, Ren XP, Chen YN, Huang L, Zhou XJ, Huang JQ, Froenicke L, Yu JJ, Guo BZ, Liao BS (2013) Phenotypic evaluation of the Chinese mini-mini core collection of peanut (*Arachis hypogaea* L.) and assessment for resistance to bacterial wilt disease caused by *Ralstonia solanacearum*. Plant Genet Resour 11:77–83

Jiang HF, Huang L, Ren XP, Chen YN, Zhou XJ, Xia YL, Huang JQ, Lei Y, Yan LY, Wan LY, Liao BS (2014) Diversity characterization and association analysis of agronomic traits in a Chinese peanut (*Arachis hypogaea* L.) mini-core collection. J Integr Plant Biol 56:159–169

Jianwei L, Huifang J, Xiaoping R, Xiaojie Z, Boshou L (2010) Identification and molecular traits of ICRISAT mini core collection of peanut species with resistance to bacterial wilt. Chin Agric Sci Bull 26:47–51

Kochert G, Halward T, Branch WD, Simpson CE (1991) RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. Theor Appl Genet 81:565–570

Kochert G, Stalker HT, Gimenes M, Galgaro L, Romero Lopes C, Moore K (1996) RFLP and cytogenetic evidence on the origin and evolution of the allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). Am J Bot 83:1282–1291

Kottapalli KR, Burow MD, Burow G, Burke J, Puppala N (2007) Molecular characterization of the U.S. peanut mini core collection using microsatellite markers. Crop Sci 47:1718–1727

Krapovickas A (1969) The origin, variability and spread of the groundnut (*Arachis hypogaea*) (English translation). In: Ucko PJ, Falk IS (eds) The domestication and exploitation of plants and animals. Gerald Duckworth Co Ltd, London, pp 424–441

Krapovickas A, Gregory WC (1994) Taxonomia del Genero *Arachis* (Leguminosae). Bonplandia 8:1–186

Liang X, Chen X, Hong Y, Liu H, Zhou G, Li S, Guo B (2009) Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and *Arachis* wild species. BMC Plant Biol 9:261–265

Liao BS (2014) Peanut Breeding. In: Nalini M, Varshney RK (eds) Genetic, genomics and breeding of peanuts. CRC Press Taylor and Francis Group, Boca Raton, London, New York, pp 61–78

Liao BS, Lei Y, Wang SY, Li D, Huang JQ, Jiang HF, Ren XP (2008) Genetic diversity of peanut RILs and enhancement for high oil genotypes. Acta Agron Sin 34:999–1004

Liao BS, Lei Y, Li D, Wang SY, Huang JQ, Ren XP, Jiang HF, Yan LY (2010) Novel high oil germplasm with resistance to *Aspergillus flavus* and bacterial wilt developed from recombinant inbred lines. Acta Agron Sin 36:1296–1301

Mallikarjuna N (2002) Gene introgression from *Arachis glabrata* into *A. hypogaea*, *A. duranensis* and *A. diogoi*. Euphytica 124:99–105

Mallikarjuna N, Jadhav DR, Reddy D, Husain F, Das K (2012) Screening new *Arachis* amphidiploids, and autotetraploids for resistance to late leaf spot by detached leaf technique. Eur J Plant Pathol 132(1):17–21

Moretzsohn MC, Hopkins MS, Mitchell SE, Kresovich S, Valls JFM, Ferreira ME (2004) Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hyber variable regions of the genome. BMC Plant Biol 4:1–10

Moretzsohn MC, Gouvea EG, Inglis PW, Leal-Bertioli SCM, Valls JFM, Bertioli DJ (2012) A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. Ann Bot 111:113–126

Nageswara Rao RC, Wright GC, Cruickshank AL (2000) Genetic enhancement of drought resistance in Australian peanuts. Proceedings American Peanut Research and Education Society 32: 71 (abstr)

Nigam SN, Chandra S, Rupa Sridevi K, Bhukta M, Reddy AGS, Rahaputi NR, Wright GC, Reddy PV, Deshmukh MP, Mathur RK, Basu MS, Vasundhara S, Vindhiyavarman P, Nagda AK (2005) Efficiency of physiological trait-based and empirical selection approaches for drought tolerance in groundnut. Ann Appl Biol 146:433–439

Nigam SN, Waliyar F, Aruna R, Reddy SV, Lava Kumar P, Craufurd PQ, Diallo AT, Ntare BR, Upadhyaya HD (2009) Breeding peanut for resistance to aflatoxin contamination at ICRISAT. Peanut Sci 36:42–49

Norden AJ, Gorbet DW, Knauft DA, Young CT (1987) Variability in oil quality among peanut genotypes in the Florida breeding program. Peanut Sci 14:7–11

Pandey MK, Monyo E, Ozias-Akins P, Liang X, Guimaraes P, Nigam SN, Upadhyaya HD, Janila P, Zhang X, Guo B, Cook DR, Bertioli DJ, Michelmore R, Varshney RK (2012) Advances in *Arachis* genomics for peanut improvement. Biotechnol Adv 30:639–651

Pittman RN (1995) United States Peanut Descriptors. USDA-ARS-132. U.S. Government Printing Office, Washington, DC

Ravi K, Vadez V, Isobe S, Mir RR, Guo Y, Nigam SN, Gowda MVC, Radhakrishnan T, Bertioli DJ, Knapp SJ, Varshney RK (2011) Identification of several small main-effect QTLs and a large number of epistatic QTLs for drought tolerance related traits in groundnut (*Arachis hypogaea* L.). Theor Appl Genet 122:1119–1132

Reddy TY, Reddy VR, Anbumozhi V (2003) Physiological responses of groundnut (*Arachis hypogaea* L.) to drought stress and its amelioration: a critical review. Plant Growth Regul 41:75–88

Seijo G, Lavia GI, Fernandez A, Krapovickas A, Ducasse DA, Bertioli DJ, Moscone EA (2007)

Genomic relationships between the cultivated peanut (*Arachis hypogaea*, Leguminosae) and its close relatives revealed by double GISH. Am J Bot 94:1963–1971

Simpson CE, Nelson SC, Starr J, Woodward KE, Smith OD (1993) Registration of T*x*AG-6 and T*x*AG-7 peanut germplasm lines. Crop Sci 33:1418

Singh AK, Mehan VK, Nigam SN (1997) Sources of resistance to groundnut fungal and bacterial wilt diseases: an update and appraisal. Information Bulletin No. 50. ICRISAT, Patancheru, AP, India, p 48

Singh AK, Dwivedi SL, Pande S, Moss JP, Nigam SN, Sastri DC (2003) Registration of rust and late leaf spot resistant peanut germplasm lines. Crop Sci 43:440–441

Stalker HT, Simpson CE (1995) Germplasm resources in *Arachis*. In: Pattee HE, Stalker HT (eds) Advances in peanut science. American Peanut Research and Education Society, Stillwater, OK, pp 14–53

Stalker HT (1997) Peanut (*Arachis hypogaea* L.). Field Crops Res 53:205–217

Stalker HT, Mozingo LG (2001) Molecular markers of *Arachis* and marker assisted selection. Peanut Sci 28:117–123

Stalker HT, Tallury SP, Ozias-Akins P, Bertioli D, Bertioli SCL (2013) The value of diploid peanut relatives for breeding and genomics. Peanut Sci 40:70–88

Subrahmanyam P, McDonald D, Waliar F, Reddy LJ, Nigam SN, Gibbons RW, Rao VR, Singh AK, Pande S, Reddy PM, Rao PVS (1995) Screening methods and sources of resistance to rust and late leaf spot of groundnut. Information Bulletin No. 47. ICRISAT, Pattancheru, AP, India, p 20

Subrahmanyam P, Naidu RA, Reddy LJ, Kumar PL, Ferguson ME (2001) Resistance to groundnut rosette disease in wild *Arachis* species. Ann Appl Biol 139:45–50

Sun DR (1998) Breeding of Groundnut. China Agricultural Press, Beijing (in Chinese). ISBN: 7-109-05154-4/S.3261

Upadhyaya HD, Ferguson ME, Bramel PJ (2001a) Status of *Arachis* germplasm collection at ICRISAT. Peanut Sci 28:89–96

Upadhyaya HD, Ortiz R, Bramel PJ, Singh S (2001b) Development of Asia region groundnut core collection. In: Diamond jubilee symposium on hundred years of post-mendelian genetics-retrospect and prospects. Indian Agricultural Research Institute, New Delhi, India, 6–9 November 2001

Upadhyaya HD, Bramel PJ, Ortiz R, Singh S (2002) Developing a mini core of peanut for utilization of genetic resources. Crop Sci 42:2150–2156

Upadhyaya HD (2003) Phenotypic diversity in groundnut (*Arachis hypogaea* L.) core collection assessed by morphological and agronomical evaluations. Genet Res Crop Evol 50:539–550

Upadhyaya HD, Ortiz R, Bramel PJ, Singh S (2003) Development of groundnut core collection using taxonomical, geographical and morphological descriptors. Genet Res Crop Evol 50:139–148

Upadhyaya HD (2005) Variability for drought resistance related traits in the mini core collection of peanut. Crop Sci 45:1432–1440

Upadhyaya HD, Reddy LJ, Gowda CLL, Singh S (2006) Identification of diverse groundnut germplasm: sources of early-maturity in a core collection. Field Crops Res 97:261–271

Upadhyaya HD, Bhattacharjee R, Hoisington DA, Chandra S, Varshney RK, Valls JFM, Moretzsohn MC, Leal-Bertioli S, Guimaraes P, Bertioli D (2008) Molecular characterization of groundnut (*Arachis hypogea* L.) composite collection. In: Project Abstracts, GCP Annual Meeting, pp 51–52. Bangkok, Thailand, 16–20 September 2008

Upadhyaya HD, Mukri G, Nadaf HL, Singh S (2012) Variability and stability analysis for nutritional traits in the mini core collection of peanut. Crop Sci 52:168–178

Upadhyaya HD, Sharma S, Dwivedi SL (2014) Genetic resources, diversity and association mapping in peanut. In: Nalini M, Varshney RK (eds) Genetic, genomics and breeding of peanuts. CRC Press Taylor and Francis Group, Boca Raton, London, New York

Valls JFM, Simpson CE (2005) New species of *Arachis* (Leguminosae) from Brazil, Paraguay and Bolivia. Bonplandia 14:35–63

Waliyar F, Vijay Krishna Kumar K, Diallo M, Traore A, Mangala UN, Upadhyaya HD, Sudini H (2016) Resistance to pre-harvest aflatoxin contamination in ICRISAT's groundnut mini core collection. Eur J Plant Pathol 145(4):901–913

Wang ML, Sukumaran S, Barkley NA, Chen Z, Chen CY, Guo B, Pittman RN, Stalker HT, Holbrook CC, Pederson GA, Yu J (2011) Population structure and marker-trait association analysis of the U.S. peanut (*Arachis hypogaea* L.) mini-core collection. Theor Appl Genet 123:1307–1317

Wang ML, Chen CY, Tonnis B, Barkley NA, Pinnow DL, Pittman RN, Davis J, Holbrook CC, Stalker HT, Pederson GA (2013) Oil, fatty acid, flavonoid, and resveratrol content variability and *FAD2A* functional SNP genotypes in the U.S. peanut mini-core collection. J Agric Food Chem 61:2875–2882

Wang HM, Lei Y, Wan LY, Yan LY, Lv JW, Dai XF, Ren XP, Guo W, Jiang HF, Liao BS (2016)

Comparative transcript profiling of resistant and susceptible peanut post-harvest seeds in response to aflatoxin production by *Aspergillus flavus*. BMC Plant Biol 16:1–16

Wright GC, Nageswara Rao RC, Farquhar GD (1994) Water-use efficiency and carbon isotope discrimination in peanut under water deficit conditions. Crop Sci 34:92–97

Yu SL (2011) Peanut Genetics and Breeding in China. Shanghai Scientific and Technology Press, Shanghai, China (in Chinese). ISBN: 978-7-5478-0610-4/S-23

# Genomics Resources for Peanut Improvement

6

Peggy Ozias-Akins, Ethalinda K.S. Cannon and Steven B. Cannon

**Abstract**

Peanut or groundnut (*Arachis hypogaea*), an allotetraploid species, recently became the focus of a global genomics initiative with the goal of developing genome-wide molecular resources to facilitate breeding. Publicly available reference genome sequence of peanut's diploid progenitors (peanutbase.org) is enabling expansion of genotyping tools for genetic mapping of quantitative trait loci (QTL) for disease resistance and reproductive traits. Gene models supported by transcriptome data from cultivated peanut and its wild relatives empower the association of allelic variants with putative function, establishing a foundation for gene-based selection. This review summarizes recent advances and emerging impacts.

## 6.1 Introduction

Cultivated peanut, *Arachis hypogaea* L., has joined the mainstream of crops for which genome sequence is being generated (Ozias-Akins 2013). Its genome size is relatively large (2.8 Gb) (Temsch and Greilhuber 2000; Samoluk et al. 2015) even in comparison with other legume crop genomes, and polyploid with highly similar A- and B-subgenomes (Bertioli et al. 2016). A comprehensive description of genomics resources for peanut (*Arachis hypogaea* L.) published in 2012 (Pandey et al. 2012b) captured the beginning of an exponential increase in genome and transcriptome sequence and molecular marker data generation for use in breeding, causing peanut to emerge from its orphan crop status (Ozias-Akins 2013; Varshney 2016). The increase in data resources has been fueled in part by the International Peanut Genome Initiative (IPGI) which led a coordinated global effort to launch the Peanut Genome Sequencing project in

P. Ozias-Akins (✉)
Institute of Plant Breeding, Genetics and Genomics, University of Georgia, Tifton, GA, USA
e-mail: pozias@uga.edu

E.K.S. Cannon
Department of Computer Science, Iowa State University, Ames, IA, USA

S.B. Cannon
Corn Insects and Crop Genetics Research Unit, US Department of Agriculture–Agricultural Research Service, Ames, IA, USA


R.K. Varshney et al. (eds.), *The Peanut Genome*, Compendium of Plant Genomes, DOI 10.1007/978-3-319-63935-2_6

2012 (Guo et al. 2013). Recognizing the complexity of the tetraploid cultivated peanut genome, priority was given to sequencing the genomes of its diploid progenitors, *A. duranensis* (AA genome) and *A. ipaensis* (BB genome), and their combined use as proxies for the tetraploid (AABB) genome (Bertioli et al. 2016). The effort to produce a reference genome sequence for the tetraploid continues and is benefitting from progressive technological improvements; however, there is considerable discovery that can be applied for breeding with the aid of diploid genome sequences. Advances in resources from 2012 until present will be the focus of this chapter.

## 6.2 Molecular Markers and Genetic Maps

### 6.2.1 Consensus Maps

Peanut breeding programs have only recently begun to make use of marker-assisted selection as marker-trait associations have been identified. Genetic maps before 2012 contained almost exclusively simple sequence repeat (SSR) markers, although these markers have been very useful for identifying quantitative trait loci (QTL) for biotic and abiotic stress tolerance and quality traits (Pandey et al. 2012b). An integrated linkage map with 324 mostly SSR markers covered 21 linkage groups and 1352 cM, and also contained two major QTL for tomato spotted wilt virus resistance, a potentially devastating disease in the USA (Qin et al. 2012). SSR markers also were used to develop an extensive consensus map based on 897 loci mapped in one backcross and 10 recombinant inbred line (RIL) populations (Gautami et al. 2012). The markers covered 20 linkage groups with a total map distance of 3863.57 cM. The RIL populations in this study encompassed different botanical types, cultivars, and breeding lines from India, China, and the USA. The backcross population was interspecific-derived, contributing uniquely 23% of the markers mapped. This high information content from an interspecific hybrid was not unexpected given the low levels of polymorphism repeatedly demonstrated within the cultivated gene pool. Another highly polymorphic marker type used for mapping (Shirasawa et al. 2012b) was derived from miniature inverted-repeat transposable elements (MITEs or AhTE markers) (Shirasawa et al. 2012a). Over 50% of the markers were polymorphic in one population, more than double the 1–20% polymorphism observed with SSR markers on these parental pairs. Further merging of diploid and tetraploid genetic maps resulted in the placement of 3693 marker loci, both genomic- and EST-SSR and AhTE, on a consensus map of the AABB genome (Shirasawa et al. 2013). The markers comprised 20 linkage groups with a total map distance of 2651 cM.

### 6.2.2 Simple Sequence Repeat (SSR) Markers

Marker-assisted breeding (MAB) in peanut is poised to expand with the development of high-throughput SNP assays and genome-wide SNP and SSR marker coverage, although low levels of polymorphism remain a limitation. SSR marker discovery has continued as additional DNA sequence, from genome and transcriptome, has been generated from diverse sets of germplasm (Table 6.1). In an effort to identify highly polymorphic SSRs, Macedo et al. (2012) enriched for long TC/AG repeats by targeted library construction and sequencing, recovering clones with microsatellites averaging 23 repeats. Of 193 repeat-containing clones, 135 were suitable for primer design and over 50% of the primer pairs amplified polymorphic loci in cultivated peanut. An average of 5.5 alleles per locus was detected, which exceeded the 3.8 average number of alleles obtained by Pandey et al. (2012a) for a selected group of 199 highly polymorphic SSR markers out of a total of 3582 collected from numerous studies. BAC (Bacterial artificial chromosome) end sequences, from clones of cultivar Tifrunner selected to contain resistance gene homologs, allowed the design of 1152 amplifiable SSRs, 12.8% of which were

polymorphic when screened with DNAs from eight genotypes and detected an average of 3.2 alleles per locus (Wang et al. 2012). The polymorphism frequency for different repeats ranged from 17.5% (AAT motif) to 4.5% (AAG motif). Dinucleotide repeat polymorphisms were intermediate in this range (14.5% for AT motif; 10.9% for AG motif). Of the 1343 polymorphic SSR markers, out of 9274 compiled by Zhao et al. (2012), the AG motif was most highly represented (36.5%) and AT the least (10.3%). This observation is consistent with the high level of polymorphism observed by Macedo et al. (2012) for SSR markers containing the AG repeat. Collections of peanut SSR markers have been assembled by several groups, although none is comprehensive. A total of 15,125 SSRs are cataloged in the Kasuza marker searchable database for *Arachis* (http://marker.kazusa.or.jp/) (Shirasawa et al. 2014); Zhao et al. (2012, Additional file 1) published a cross-referenced list of 980 polymorphic SSRs; and Guo et al. (2012) also cataloged, renamed, and cross-referenced previously published SSRs and developed more than 2000 new EST-SSRs (Guo et al. 2012, Additional files 2 and 7). In addition to SSRs published prior to 2012, the Kasuza database includes the Macedo et al. (2012) long-repeat SSRs and Wang et al. (2012) BAC-end sequence SSRs, some of the latter being redundant with previously published sequences, but nevertheless providing different primer options. Re-mining new assemblies from publicly available data sets also has contributed to redundancy across studies. Several groups have continued to mine transcriptome data for SSRs (Zhang et al. 2012; Bosamia et al. 2015; Zhong et al. 2016; Peng et al. 2016), but only two searched their assemblies for previously published SSRs (Bosamia et al. 2015; Peng et al. 2016). In Peng et al. (2016), both previously published and newly generated transcriptome sequence yielded over 7500 SSRs, 6455 of which were considered to be novel and for which primers were designed. Of these, 380 markers were selected for validation, of which 89% amplified, although only 12% detected polymorphism among four cultivars. Similar results were obtained by Bosamia et al. (2015) where 2456 novel EST-SSR primer pairs were designed, 366 were tested, and of the 340 that yielded clear amplification products, 10.66% were polymorphic. Peanut gynophore (Zhong et al. 2016) and seed (Zhang et al. 2012) transcriptome assemblies yielded 5,058 and 5,883 SSRs of which 200 and 160, respectively, were tested for amplification and polymorphism yielding 8% (across 16 cultivars) and 40% (across six varieties) that were polymorphic.

### 6.2.3 Single Nucleotide Polymorphism (SNP) Markers

The high degree of colinearity between diploid progenitor and tetraploid subgenomes shown from genetic mapping studies supported the choice of diploid ancestral lines, as precedent to the cultivated species, for genome sequencing (Bertioli et al. 2016). The B-genome progenitor, *A. ipaensis*, is nearly identical to the B-subgenome of *A. hypogaea* and has limited geographical distribution, having been collected from a single site. Conversely, extensive polymorphism between accessions of the A-genome progenitor, *A. duranensis*, guided marker discovery and construction of a dense genetic map consisting of 1054 single nucleotide polymorphism (SNP) and 598 EST-SSR markers (Nagy et al. 2012). SNP markers are more abundant in genomes relative to SSRs, amenable for haplotyping, and demonstrate less frequent homoplasy (Rafalski 2002; Batley and Edwards 2007; Ganal et al. 2009); however, high-throughput assays for SNPs in polyploids present unique challenges. These challenges have been outlined in Clevenger et al. (2015a) and are particularly acute for peanut where the subgenomes diverged from one another only approximately 2 million years ago (Bertioli et al. 2016). An improvement in SNP detection accuracy for peanut was achieved by implementing a SNP calling pipeline that takes flanking SNPs, or haplotypes, into account when comparing sequence variation within vs. between genotypes (Clevenger and Ozias-Akins 2015). With

**Table 6.1** Marker resources published since Pandey et al. (2012b)

| Marker type | Marker designation | Number of primer pairs/assays designed | Number tested | Number/% polymorphic in cultivated (number lines screened) | References |
|---|---|---|---|---|---|
| MITE | AhTE | 504 | 504 | 169/33% (4) | Shirasawa et al. (2012a) |
| Genomic SSR | AHGS | 6680 | 2167 | 675/31% (4) | Shirasawa et al. (2012b) |
| MITE | AhTE | 535 | 535 | 304/57% (4) | Shirasawa et al. (2012b) |
| Genomic SSR (BES) | Ad or Ai | 152 | 152 | 25/16% (2) | Shirasawa et al. (2012b) |
| Genomic SSR | TC | 146 | 146 | 78/53% (22) | Macedo et al. (2012) |
| Genomic SSR (BES) | GNB | 1152 | 1152 | 148/13% (8) | Wang et al. (2012) |
| EST-SSR | DGR | 2456 | 366 | 39/11% (11) | Bosamia et al. (2015) |
| EST-SSR | GM | 2138 | 2138 | 203/9% (4) | Guo et al. (2012) |
| EST-SSR | AHW | 6455 | 380 | 22/6% (4) | Peng et al. (2016) |
| EST-SSR | HAS | 3919 | 160 | 65/41% (6) | Zhang et al. (2012) |
| EST-SSR | GU | 200 | 200 | 17/9% (16) | Zhong et al. (2016) |
| SNP | GKAM | 96 | 96 | 64/67% (68) | Khera et al. (2013) |
| SNP | EST- or GSS-nnnSNP | 34 | 34 | 33/97% (96) | Hong et al. (2015) |

this strategy, true SNPs are more easily distinguished from homeologous SNPs. Using tetraploid resequencing data, the pipeline was used to identify SNPs for inclusion on a 60K Affymetrix array (Clevenger et al. 2017; Pandey et al. 2017).

Tetraploid-derived SNP markers were first identified computationally from transcriptome sequence of 19 genotypes and validated with a GoldenGate array and as KASP (Kompetitive Allele Specific PCR) markers (Khera et al. 2013). The interference of duplicated sequences with SNP calling and the high frequency of detecting alleles in both subgenomes became apparent. Manual correction of signal intensity clusters was necessary since clusters were distorted in the tetraploid due to homeolog signal; nevertheless, ~25% of the SNPs on the GoldenGate array were informative in an interspecific tetraploid population (Bertioli et al. 2014). Smaller scale validation and application of SNPs confirmed the robustness of KASP assays for genotyping in peanut (Chopra et al. 2015). An alternative SNP assay, high-resolution melting also has been used to validate SNPs in peanut (Hong et al. 2015). In this study, 12 tetraploid varieties formed a panel for PCR amplification and Sanger sequencing of 200 EST or genomic regions, and SNPs in 30 of them were successfully converted to high-resolution melting assays. Methods of sequence-based genotyping (Poland and Rife 2012) provide an alternative to SNP arrays for detecting variants, although the same homeolog interference (apparent heterozygosity) will be encountered during analysis of sequence data. A ddRADseq approach was taken by Zhou et al. (2014) in which DNAs from two parents and 166 RILs from a cross of these two were subjected to double digestion with restriction enzymes, *Sac*I and *Mse*I, in order to reduce the fraction of the genome sampled. Stringent data filtering to remove SNPs scored as heterozygous within a parent (homeolog detection since these are inbred lines), supported by fewer than four reads, or missing data in >25% of the population resulted in the recovery of 1765 SNPs, 1621 of which could be mapped onto 20 linkage groups. Gupta

et al. (2015) also used a ddRADseq approach with different restriction enzymes, *Hin*dIII and *Nla*III, screened seven cultivated peanut genotypes, and recovered 3248 SNPs after filtering, but did not conduct any SNP validation. The relatively low yield of SNPs from sequence-based genotyping in peanut compared with other crops further illustrates the computational challenges and low genetic diversity within the crop.

## 6.3 Marker-Trait Associations for Marker-Assisted Selection

### 6.3.1 Marker-Assisted Selection

Relatively few traits were amenable to marker-assisted selection prior to 2010. One notable exception is the trait for high oleic to linoleic acid ratio (high O/L). This seed oil composition is in demand by the peanut industry because it reduces the oxidative potential of oil in stored peanuts and slows the development of rancidity. It also has health benefits, thus is attractive to the consumer. Extensive characterization of oil biosynthetic pathways in other oilseed crops simplified the isolation and sequence analysis of mutant alleles of fatty acid desaturase 2 (FAD2) in peanut. Mutations in each duplicate gene of the tetraploid, are necessary to achieve high O/L peanut by reducing the conversion of oleic to linoleic acid. The A-genome mutation is a SNP whereas three B-genome mutant alleles have been described, a single nucleotide or MITE indel (insertion–deletion) (Jung et al. 2000; Lopez et al. 2000; Patel et al. 2004) or a single nucleotide substitution (Wang et al. 2015b). Markers have been designed to rapidly screen for these functional mutations at the individual seed or seedling stage (Chu et al. 2007, 2009; Barkley et al. 2011b; Chen et al. 2010), and have been applied in multiple breeding programs (Chu et al. 2011; Barkley et al. 2011a, 2013; Janila et al. 2016). Even for this well-characterized, mainly two-gene-controlled trait, marker design can be problematic because of the high level of DNA sequence similarity between the A- and B-genome copies of FAD2. Runner breeding programs benefit from a high frequency of the mutant A-genome allele in ssp. *hypogaea* (Chu et al. 2007) and often can treat the trait as single gene inheritance when the mutant allele is fixed in breeding lines. Breeding programs focused on ssp. *fastigiata* germplasm typically will encounter wild-type alleles in both A- and B-genome copies (Chu et al. 2007; Mukri et al. 2012; Wang et al. 2011, 2013b). The FAD2 genes have been mapped to linkage groups A09 and B09 in cultivated peanut (Qin et al. 2012), although other QTL affecting oleic or linoleic acid content and O/L ratio were identified in other regions of the genome (Pandey et al. 2014b). QTL corresponding to the region where FAD2 genes are located were consistently detected, and the B-genome copy explained more of the phenotypic variation than the A-genome copy when both mutant alleles were segregating in a population. Other oil biosynthesis gene family members (Shilman et al. 2011; Wang et al. 2015a) and QTL for minor fatty acids (Wang et al. 2015c) have been characterized in peanut, and this knowledge will contribute to breeding for a higher quality crop.

Another trait for which marker-assisted breeding (MAB) was applied early on is root-knot nematode (RKN) resistance. This trait was introgressed into cultivated peanut from an A-genome wild relative, *A. cardenasii* (Simpson 2001); hence, polymorphic markers in the introgressed region from chromosome A09 of *A. cardenasii* are relatively abundant (Nagy et al. 2010). MAB for nematode resistance has been successful (Chu et al. 2011) but plagued with the eventual consequences of low recombination and potential linkage drag. With rare recombination, the distance of a marker used for selection from the gene(s) conferring the trait may remain unknown, and if large, has a higher probability that linkage will be broken over many generations. This has indeed been demonstrated for RKN resistance in peanut (Branch et al. 2014) and rare recombinants have allowed refinement of marker selections for MAB (Chu et al. 2016).

A more recent example of MAB is for introgression of a major QTL for rust resistance

derived from interspecific hybrid materials (Varshney et al. 2014). The QTL explained up to 82% of the phenotypic variation observed in segregating populations from a cross involving a rust-resistant parent. One dominant and three codominant SSR markers were used to select for $BC_2F_2$ and $BC_3F_2$ lines homozygous for the resistance marker alleles, and these lines demonstrated resistance levels equivalent to the resistant parent when tested in the field. The same markers differentiated near isogenic lines extracted from related materials; isogenic lines with resistant parent alleles were resistant and those with susceptible parent alleles were susceptible (Yeri et al. 2014). Additional markers for rust resistance that potentially can be used for MAB have been identified by Mondal et al. (2014b), who utilized 243 of the AhTE markers from Shirasawa et al. (2012b) to screen resistant and susceptible parents and population bulks. Of the 16% that were polymorphic between the two parents, two were polymorphic between the bulks and were associated with rust resistance in the segregating RIL population, explaining 65–75% of the phenotypic variation. Although the source of rust resistance is reportedly different, the AhTE markers are linked with one of the markers (IPAHM103) used for MAS by Varshney et al. (2014) and Yeri et al. (2014). An alternative source of rust resistance that could be useful for cultivated peanut upon introgression is derived from the wild relative, *A. magna*. QTL for rust resistance were identified in a cross between two B-genome diploids, *A. ipaensis* x *A. magna*. A major QTL for multiple resistance components (time of appearance of the first lesion, number of lesions per leaf area, number of sporulated lesions per leaf area, and sporulation index) that explained 5.8–59.3% of the phenotypic variation was identified on B08. Twenty-two KASP assays were designed and validated on diploid and tetraploid materials, providing reagents that will facilitate introgression of this new resistance source (Leal-Bertioli et al. 2015a).

### 6.3.2 QTL Mapping

Additional QTL mapping studies have broadened our understanding of the complexity of pest and disease resistance, plant growth habit, and drought tolerance (Table 6.2). These studies extend to wild species and introgression lines incorporating wild species germplasm. The most extensively studied wild species to date is *A. duranensis*, one of the diploid progenitors of cultivated peanut for which considerable intraspecific diversity exists. Aside from the dense molecular map published for *A. duranensis* (Nagy et al. 2012), introgressed QTL have been identified in a cultivated background upon analysis of chromosome segment substitution lines (CSSLs; Fonceka et al. 2009, 2012a, b). Of the 42 QTL identified, 14 were for plant growth habit (7 on 9 A-genome linkage groups), 15 for height of the main stem (6 on 5 A-genome linkage groups), 12 for plant spread (6 on 5 A-genome linkage groups), and one for flower color (A-genome). These QTL were defined by the sizes of introgressed segments. Further analysis of this same population under water-sufficient and water-limited conditions identified 95 QTL, 53 on A-genome linkage groups, for days to flowering, plant architecture, pod morphology, seed morphology, yield components, and stress tolerance. Given the differences in plant morphology between wild diploids and cultivated tetraploids, identification of QTL for drought-related and other traits in tetraploid introgression lines is likely to be more informative than direct evaluation of such traits in diploid mapping populations (Leal-Bertioli et al. 2012). Variation for drought tolerance exists within the cultivated germplasm and the genetic basis is being further explored through genetic mapping. Extending from a previous study on drought tolerance in a RIL population of TAG 24 × ICGV 86031 (Ravi et al. 2011), Faye et al. (2015) discovered 52 QTL upon screening for yield and yield component traits in a west-African environment. Few of these QTL co-localized with those identified in Ravi et al. (2011) whose study was conducted in India. Environment-specific effects were observed

when testing 247 lines from the ICRISAT groundnut reference collection under well-watered and water-stressed conditions in India and Niger (Hamidou et al. 2012).

Disease resistance QTL identified in wild diploids are more likely to persist in introgression lines compared with developmental or drought tolerance traits. The wild diploids provide a rich resource of disease resistance alleles, some of which underlie QTL identified through genetic mapping studies. An interspecific RIL population between two B-genome species, *A. ipaensis* and *A. magna*, revealed 13 QTL for rust resistance, only three of which were from the rust-resistant parent, *A. magna*, although these three had the biggest effect, explaining 33–59% of the phenotypic variation (Leal-Bertioli et al. 2015a). An A-genome interspecific RIL population (*A. duranensis* x *A. stenosperma*) allowed discovery of eight QTL for nematode resistance, all but one derived from the resistant parent, *A. stenosperma* (Leal-Bertioli et al. 2015b). Each of these disease resistance studies also mapped morphological and productivity traits including seed number, seed weight, peg length, pod constriction, main stem height, number and length of lateral branches, aerial and root dry weight, ratio of root to aerial weight, total biomass, pollen viability, and drought-related traits such as SPAD chlorophyll meter reading and specific leaf area. Knowledge of these QTL may also be useful for monitoring linkage drag in crosses with tetraploids.

Although the strongest disease resistance alleles are found in the wild relatives of peanut, their transfer to the cultivated types is long-term since it involves overcoming interspecific hybridization barriers and linkage drag. There remain underutilized and untapped sources of disease resistance in the cultivated species. Genetic mapping of these resistance genes/alleles can facilitate their future use in elite breeding materials. Peanut suffers from several viral diseases, one of which almost devastated peanut cultivation in the southeastern USA when a monoculture of tomato spotted wilt virus-susceptible Florunner was practiced (Holbrook and Stalker 2003). A source of resistance, PI 203396, was quickly identified and incorporated into breeding programs, one of which gave rise to Tifrunner (Holbrook and Culbreath 2007). The genetic basis of resistance is unknown, however. QTL mapping of populations with Tifrunner as a parent has identified 9–15 QTL the largest of which explained up to 35% of phenotypic variance (Wang et al. 2013a). The genetic basis of resistance to TSWV in other germplasm is being studied to ensure allelic diversity.

Increasing yield is a primary goal for cultivar improvement and some peanut uses require haulm as well as pod yield. QTL have been identified for the production of above and below ground biomass, harvest index as an expression of this ratio, and seeds or pods per plant (Table 6.2) (Faye et al. 2015; Fonceka et al. 2012a, b; Shirasawa et al. 2012b). Other components of yield such as pod and seed size or weight have been extensively studied, often, but not always, in crosses between divergent market types (Chen et al. 2016; Fonceka et al. 2012a, b; Huang et al. 2015; Shirasawa et al. 2012b).

In addition to yield, quality traits are of utmost importance in peanut. The most extensive analyses have been carried out on oil content and fatty acid composition traits (Table 6.2). One trait considered desirable by the peanut industry is a high ratio of oleic to linoleic acid. While several QTL for this trait were identified by Pandey et al. (2014b), the most consistent QTL overlapped with the known locations of AhFAD2 genes in which functional mutations in this enzyme of the biosynthetic pathway controlling O/L ratio have been identified and are being used for marker-assisted selection (see above). The single QTL detected for oleic acid by Huang et al. (2015) was not on the same linkage group as AhFAD2, suggesting that modifier loci may be involved in altering oleic acid content. Other fatty acids contribute to peanut's oil properties and multiple QTL affecting their levels have been identified (Huang et al. 2015; Wang et al. 2015c). QTL for secondary metabolites/bioactive compounds also have been described (Mondal et al. 2015).

**Table 6.2** QTL mapping conducted since Pandey et al. (2012b)

| Trait | QTL number | Phenotypic variance explained (%) | Reference | Mapping population |
|---|---|---|---|---|
| **Disease/pest resistance** | | | | |
| Rust resistance | 1 | 65–75 | Mondal et al. (2014b) | VG 9514 (virginia) × TAG24 (spanish) |
| Rust resistance | 13 | 5.8–59.3 | Leal-Bertioli et al. (2015a) | (*A. ipaënsis* K 30076) × *A. magna* (GKSSc 30097) |
| Leaf spot | 37 | 6.61–27.35 | Wang et al. (2013a) | Tifrunner (runner) × GT-C20 (spanish) F2 |
| Leaf spot | 13 | 5.95–21.45 | Wang et al. (2013a) | Tifrunner (runner) × GT-C20 (spanish) F5 |
| TSWV resistance | 15 | 2.51–5.92 | Wang et al. (2013a) | Tifrunner (runner) × GT-C20 (spanish) F2 |
| TSWV resistance | 9 | 5.2–14.14 | Wang et al. (2013a) | Tifrunner (runner) × GT-C20 (spanish) F5 |
| TSWV resistance | 1 | 12.9 | Qin et al. (2012) | Tifrunner (runner) × GT-C20 (spanish) RIL |
| TSWV resistance | 1 | 35.8 | Qin et al. (2012) | SunOleic 97R (runner) × NC94022 (hirsuta-der) |
| Thrips resistance | 2 | 12.14–19.43 | Wang et al. (2013a) | Tifrunner (runner) × GT-C20 (spanish) F2 |
| Thrips resistance | 1 | 5.86 | Wang et al. (2013a) | Tifrunner (runner) × GT-C20 (spanish) F5 |
| Bruchid adult emergence | 2[a] | 24–29 | Mondal et al. (2014a) | VG 9514 (virginia) × TAG24 (spanish) |
| Bruchid development period | 4[a] | 36–67 | Mondal et al. (2014a) | VG 9514 (virginia) × TAG24 (spanish) |
| Pod damage by bruchids | 3[a] | 32–37 | Mondal et al. (2014a) | VG 9514 (virginia) × TAG24 (spanish) |
| Pod weight loss due to bruchid | 1[a] | 33 | Mondal et al. (2014a) | VG 9514 (virginia) × TAG24 (spanish) |
| **Plant morphology** | | | | |
| Plant architecture | 11 | 9.8–26.0 | Fonceka et al. (2012a) | Fleur11 (spanish) × [*A. ipaensis* K30076 x *A. duranensis* V14167]4x |
| Plant growth habit | 14 | na | Fonceka et al. (2012b) | Fleur11 (spanish) × [*A. ipaensis* K30076 x *A. duranensis* V14167]4x |
| Branch angle | 2 | 11.9–23.2 | Shirasawa et al. (2012b) | Satonoka (virginia) × Kintoki (spanish) |
| Mainstem height | 15 | na | Fonceka et al. (2012b) | Fleur11 (spanish) × [*A. ipaensis* K30076 x *A. duranensis* V14167]4x |
| Mainstem height | 3 | 6.12–8.9 | Huang et al. (2015) | Zhonghua 10 × ICG12625 |
| Mainstem height | 3 | 4.8–19.2 | Shirasawa et al. (2012b) | Satonoka (virginia) × Kintoki (spanish) |

(continued)

**Table 6.2** (continued)

| Trait | QTL number | Phenotypic variance explained (%) | Reference | Mapping population |
|---|---|---|---|---|
| Plant height | 5 | 4.04–8.16 | Faye et al. (2015) | TAG 24 (spanish) × ICGV 86031 (spanish) |
| Plant spread | 12 | na | Fonceka et al. (2012b) | Fleur11 (spanish) × [*A. ipaensis* K30076 x *A. duranensis* V14167]4x |
| Length of longest branch | 2 | 14.2–21.1 | Shirasawa et al. (2012b) | Satonoka (virginia) × Kintoki (spanish) |
| Number of primary branches | 5 | 0.04–8.58 | Faye et al. (2015) | TAG 24 (spanish) × ICGV 86031 (spanish) |
| Total branch number | 2 | 6.11–7.46 | Huang et al. (2015) | Zhonghua 10 × ICG12625 |
| Total branch number | 1 | 15.6 | Shirasawa et al. (2012b) | Satonoka (virginia) × Kintoki (spanish) |
| Flower color | 1 | na | Fonceka et al. (2012b) | Fleur11 (spanish) × [*A. ipaensis* K30076 x *A. duranensis* V14167]4x |
| **Reproductive phenology** | | | | |
| Days to flowering | 1 | 9.3 | Fonceka et al. (2012a) | Fleur11 (spanish) × [*A. ipaensis* K30076 x *A. duranensis* V14167]4x |
| Days to flowering | 1 | 19.5 | Shirasawa et al. (2012b) | Satonoka (virginia) × Kintoki (spanish) |
| **Yield components** | | | | |
| Plant weight | 1 | 11.8 | Shirasawa et al. (2012b) | Satonoka (virginia) × Kintoki (spanish) |
| Haulm yield | 9 | 3.74–10 | Faye et al. (2015) | TAG 24 (spanish) × ICGV 86031 (spanish) |
| Harvest index | 4 | 3.5–8.28 | Faye et al. (2015) | TAG 24 (spanish) × ICGV 86031 (spanish) |
| Yield components | 26 | 9.2–20.6 | Fonceka et al. (2012a) | Fleur11 (spanish) × [*A. ipaensis* K30076 x *A. duranensis* V14167]4x |
| Pod yield | 6 | 4.27–11.38 | Faye et al. (2015) | TAG 24 (spanish) × ICGV 86031 (spanish) |
| Pods per plant | 1 | 28.1 | Shirasawa et al. (2012b) | Satonoka (virginia) × Kintoki (spanish) |
| Seeds per plant | 1 | 6.8 | Shirasawa et al. (2012b) | Satonoka (virginia) × Kintoki (spanish) |
| Shelling % | 2 | 5.74–6.97 | Faye et al. (2015) | TAG 24 (spanish) × ICGV 86031 (spanish) |
| Shelling % | 3 | 2–11.78 | Huang et al. (2015) | |
| **Stress response** | | | | |
| Stress tolerance | 13 | 10.4–20.1 | Fonceka et al. (2012a) | Fleur11 (spanish) × [*A. ipaensis* K30076 x *A. duranensis* V14167]4x |
| SPAD chlorophyll meter reading | 12 | 2.96–10.4 | Faye et al. (2015) | TAG 24 (spanish) × ICGV 86031 (spanish) |

(continued)

**Table 6.2** (continued)

| Trait | QTL number | Phenotypic variance explained (%) | Reference | Mapping population |
|---|---|---|---|---|
| **Pod characteristics** | | | | |
| Pod morphology | 31 | 8.5–23.9 | Fonceka et al. (2012a) | Fleur11 (spanish) × [*A. ipaensis* K30076 × *A. duranensis* V14167] 4x |
| Pod constriction | 2 | 6.9–18.1 | Shirasawa et al. (2012b) | Satonoka (virginia) × Kintoki (spanish) |
| Pod tip shape | 1 | 9.9 | Shirasawa et al. (2012b) | Satonoka (virginia) × Kintoki (spanish) |
| Pod thickness | 1 | 21.7 | Shirasawa et al. (2012b) | Satonoka (virginia) × Kintoki (spanish) |
| Pod length | 1 | 11.23 | Huang et al. (2015) | Zhonghua 10 × ICG12625 |
| Pod length | 6 | 5.7–24.29 | Chen et al. (2016) | Fuchuan Dahuasheng × ICG 6375 |
| Pod length | 4 | 1.25–7.79 | Chen et al. (2016) | Xuhua 13 × Zhonghua 6 |
| Pod length | 3 | 8.4–28.2 | Shirasawa et al. (2012b) | Satonoka (virginia) × Kintoki (spanish) |
| Pod width | 2 | 2.11–18.7 | Huang et al. (2015) | Zhonghua 10 × ICG12625 |
| Pod width | 8 | 5.16–16.14 | Chen et al. (2016) | Fuchuan Dahuasheng × ICG 6375 |
| Pod width | 4 | 4.48–8.78 | Chen et al. (2016) | Xuhua 13 × Zhonghua 6 |
| Pod width | 2 | 15.2–25.5 | Shirasawa et al. (2012b) | Satonoka (virginia) × Kintoki (spanish) |
| 100 pod weight | 3 | 8.02–15.39 | Huang et al. (2015) | Zhonghua 10 × ICG12625 |
| **Seed morphology** | | | | |
| Sound mature kernel % | 6 | 3.3–7.41 | Faye et al. (2015) | TAG 24 (spanish) × ICGV 86031 (spanish) |
| 100 seed weight | 2 | 8.78–11.56 | Faye et al. (2015) | TAG 24 (spanish) × ICGV 86031 (spanish) |
| 100 seed weight | 3 | 1.69–17.88 | Huang et al. (2015) | Zhonghua 10 × ICG12625 |
| Seed weight | 1 | 19.1 | Shirasawa et al. (2012b) | Satonoka (virginia) × Kintoki (spanish) |
| Seed length | 3 | 9.86–10.48 | Huang et al. (2015) | Zhonghua 10 × ICG12625 |
| Seed length | 8 | 5.66–20.8 | Chen et al. (2016) | Fuchuan Dahuasheng × ICG 6375 |
| Seed length | 2 | 3.03–4.87 | Chen et al. (2016) | Xuhua 13 × Zhonghua 6 |
| Seed width | 4 | 6.39–12.2 | Huang et al. (2015) | Zhonghua 10 × ICG12625 |
| Seed width | 4 | 7.42–14.43 | Chen et al. (2016) | Fuchuan Dahuasheng × ICG 6375 |
| Seed width | 3 | 3.77–9.76 | Chen et al. (2016) | Xuhua 13 × Zhonghua 6 |
| Seed morphology | 13 | 8.7–23 | Fonceka et al. (2012a) | Fleur11 (spanish) × [*A. ipaensis* K30076 × *A. duranensis* V14167] 4x |

(continued)

**Table 6.2** (continued)

| Trait | QTL number | Phenotypic variance explained (%) | Reference | Mapping population |
|---|---|---|---|---|
| Seed coat color | 1 | 9.7 | Shirasawa et al. (2012b) | Satonoka (virginia) × Kintoki (spanish) |
| **Seed composition** | | | | |
| Oil content | 1 | 14.36 | Huang et al. (2015) | Zhonghua 10 × ICG12625 |
| Oil content | 6 | 3.07–10.23 | Pandey et al. (2014b) | SunOleic 97R (runner) × NC94022 (hirsuta) |
| Oil content | 9 | 3.93–14.07 | Pandey et al. (2014b) | Tifrunner (runner) × GT-C20 (spanish) |
| Palmitic acid | 1 | 17.02 | Huang et al. (2015) | Zhonghua 10 × ICG12625 |
| Palmitic acid | 11[a] | 1.7–22.04 | Wang et al. (2015c) | SunOleic 97R (runner) × NC94022 (hirsuta) |
| Palmitic acid | 19[a] | 3.06–37.37 | Wang et al. (2015c) | Tifrunner (runner) × GT-C20 (spanish) |
| Oleic acid | 1 | 1.72 | Huang et al. (2015) | Zhonghua 10 × ICG12625 |
| Oleic acid | 8 | 1.59–27.54 | Pandey et al. (2014b) | SunOleic 97R (runner) × NC94022 (hirsuta) |
| Oleic acid | 9 | 3.63–28.98 | Pandey et al. (2014b) | Tifrunner (runner) × GT-C20 (spanish) |
| Linoleic | 7 | 1.46–28.22 | Pandey et al. (2014b) | SunOleic 97R (runner) × NC94022 (hirsuta) |
| Linoleic | 9 | 3.91–25.49 | Pandey et al. (2014b) | Tifrunner (runner) × GT-C20 (spanish) |
| O/L Ratio | 6 | 1.04–42.33 | Pandey et al. (2014b) | SunOleic 97R (runner) × NC94022 (hirsuta) |
| O/L Ratio | 5 | 5.7–14.9 | Pandey et al. (2014b) | Tifrunner (runner) × GT-C20 (spanish) |
| Eicosanoic acid | 3 | 3.8–7.51 | Huang et al. (2015) | Zhonghua 10 × ICG12625 |
| Gadoleic | 7[a] | 2.55–8.77 | Wang et al. (2015c) | SunOleic 97R (runner) × NC94022 (hirsuta) |
| Gadoleic | 20[a] | 2.98–15.11 | Wang et al. (2015c) | Tifrunner (runner) × GT-C20 (spanish) |
| Behenic acid | 2 | 15.76–18.85 | Huang et al. (2015) | Zhonghua 10 × ICG12625 (*aequatoriana*) |
| Behenic acid | 5[a] | 2.88–6.95 | Wang et al. (2015c) | SunOleic 97R (runner) × NC94022 (*hirsuta*) |
| Behenic acid | 16[a] | 4.74–13.56 | Wang et al. (2015c) | Tifrunner (runner) × GT-C20 (spanish) |
| Lignoceric acid | 5[a] | 2.89–6.58 | Wang et al. (2015c) | SunOleic 97R (runner) × NC94022 (*hirsuta*) |

(continued)

**Table 6.2** (continued)

| Trait | QTL number | Phenotypic variance explained (%) | Reference | Mapping population |
|---|---|---|---|---|
| Lignoceric acid | 13[a] | 3.85–12.61 | Wang et al. (2015c) | Tifrunner (runner) × GT-C20 (spanish) |
| Arachidic acid | 2 | 8.1–20.2 | Huang et al. (2015) | Zhonghua 10 × ICG12625 |
| Arachidic acid | 6[a] | 3.6–6.4 | Wang et al. (2015c) | SunOleic 97R (runner) × NC94022 (*hirsuta*) |
| Arachidic acid | 14[a] | 3.05–36.93 | Wang et al. (2015c) | Tifrunner (runner) × GT-C20 (spanish) |
| Stearic acid | 2 | 2.52–18.31 | Huang et al. (2015) | Zhonghua 10 × ICG12625 |
| Stearic acid | 6[a] | 3.26–5.9 | Wang et al. (2015c) | SunOleic 97R (runner) × NC94022 (*hirsuta*) |
| Stearic acid | 15[a] | 2.63–40.57 | Wang et al. (2015c) | Tifrunner (runner) × GT-C20 (spanish) |
| Total phenolics | 1 | 12.5 | Mondal et al. (2015) | VG 9514 (virginia) × TAG24 (spanish) |
| Total flavonoids | 5 | 25–67 | Mondal et al. (2015) | VG 9514 (virginia) × TAG24 (spanish) |
| DPPH radical scavenging | 4 | 11.5–33 | Mondal et al. (2015) | VG 9514 (virginia) × TAG24 (spanish) |

[a]Multiple QTL detection software were tested but only results of QTL Cartographer are presented

### 6.3.3 Association Mapping

Association mapping and marker-trait association also provide valuable information for trait resources, particularly since it captures historical recombination events and allelic diversity not present in structured populations. Marker-trait associations have been most extensively characterized in the ICRISAT reference collection of 300 diverse genotypes (Pandey et al. 2014a). Thirty-six traits were evaluated including disease resistance (*Aspergillus*, early leaf spot, late leaf spot, groundnut rosette), plant morphology (leaf length, specific leaf area, total leaf area, leaf weight, root volume), yield components (shoot weight, haulm weight, harvest index, pod yield, shelling percentage), pod and seed characteristics (seed length, seed width, seed weight), and several of these traits under water-deficit stress. The Chinese minicore of 298 accessions also has been genotyped with SSRs and phenotyped for plant morphology (main stem height, length of first branches, total branch number, leaflet length, leaflet width, number of reproductive branches), yield components (pods per plant, pod weight per plant, shelling percentage), pod and seed characteristics (pod length, pod width, 100-pod weight, seed length, seed width, 100-seed weight), and marker-trait associations were found for all traits (Jiang et al. 2014).

### 6.3.4 Wild Species Alleles

More extensive utilization of wild relatives of peanut is anticipated for the future, given the rich source of alleles they provide (Bertioli et al. 2011; Stalker et al. 2013), improved understanding of species relationships (Robledo and Seijo 2010; Leal-Bertioli et al. 2015c), and refined introgression strategies (Simpson 2001). While disease resistance has been the trait most

studied in wild species (Leal-Bertioli 2015a, b), they also harbor novel and potentially useful alleles for abiotic stress tolerance (Fonceka et al. 2012a; Leal-Bertioli et al. 2012) and oil content (Huang et al. 2012). Wild species are not only valuable resources for breeding but also require long-term effort and consistent support. The recent explosion of molecular tools that can be applied to peanut will facilitate development of pre-breeding materials (Sharma et al. 2013) such as backcross, CSSL or AB-QTL populations (Fonceka et al. 2009, 2012a, b; Falleiro Suassuna et al. 2015) that can be phenotyped, selected for desirable traits, and introduced into breeding programs.

### 6.3.5 Gene Expression Resources

As of March 1, 2012, there were only 252,832 *Arachis* ESTs in NCBI's GenBank, 178,490 from *A. hypogaea* (Feng et al. 2012). Extensive sequencing of expressed genes has occurred over the last 4 years as costs for next-generation sequencing have dropped precipitously. Much of this sequence has been deposited in public databases along with previously generated microarray data (Table 6.3). Analyses of differential gene expression and pathways across developmental series and in response to pests, diseases, and abiotic stress have been the primary objectives across these studies. Several microarray experiments preceded high-throughput sequencing and were informative for the study of drought stress response. Expressed sequence from all primary organs of cultivated peanut is available and can be used to construct gene networks that will inform the study of traits that impact productivity (Clevenger et al. 2016). For example, an oil metabolic network of peanut and three other oilseed crops (soybean, rapeseed, and sesame) has been constructed using available EST data and is accessible at http://ocri-genomics.org/ocsESTdb (Ke et al. 2015). Knowledge of gene expression, association of expressed genes with gene models in the newly released diploid *Arachis* genomes (Bertioli et al. 2016), and putative positions with respect to QTL are contributing to hypothesis development and testing of gene function.

Testing of gene function often includes quantitative analysis of expression patterns across time, developmental stages, treatments, etc. Such analyses require a frame of reference. i.e., genes whose expression is relatively uniform regardless of tissue or treatment. Four studies to identify suitable reference genes for peanut have been conducted, testing 8–14 candidates across a variety of tissues (Table 6.4). Both alcohol dehydrogenase III and alpha tubulin showed high stability and moderate abundance in more than one study.

In addition to extensive transcriptome data, proteome data are being produced to further evaluate gene function, study physiological responses, and develop proteome maps. While a couple of studies have been focused on the peanut leaf proteome, with or without stress (Kottapalli et al. 2009; Katam et al. 2010), most have targeted peanut gynophores (Li et al. 2013; Sun et al. 2013; Xia et al. 2015, Zhao et al. 2015b), organs that are relatively unique among crop plants but whose response to gravity is critical for reproductive success. In addition to being transcribed, most genes must be translated to perform a function, and the coordination of transcription and translation can be different for different genes depending on the regulation of expression and turnover rates. Analyzing the proteome independently and in concert with the transcriptome can provide new insight into gene action. For example, Zhao et al. (2015b) only identified transcripts for 38 out of 69 key peg proteins while transcripts were not identified for some abundant proteins and vice versa.

## 6.4 Bioinformatics Resources

Bioinformatics resources for peanut at the time of writing include the genome sequences of the diploid progenitors of cultivated peanut, *A. duranensis* and *A. ipaensis*, as well as extensive transcriptome sequences for wild diploid accessions and for cultivated peanut (Bertioli et al. 2016). The genomic sequence and assembly is available at

**Table 6.3** Publicly available gene expression data sets

| Data type | NCBI accession | Publication | Target tissues | Treatment/ Experimental materials[a] | Objective |
|---|---|---|---|---|---|
| RNAseq | PRJNA312741 | Qingdao Agric Univ, unpubl. | Leaf | Salinity stress | |
| RNAseq | PRJNA298453 | Shandong Acad Agric Sci, unpubl. | Gynophore | Light and dark | |
| RNAseq | PRJNA291488 | Bertioli et al. (2016) | 22 | Developmental stages | Also accessible at peanutbase.org |
| RNAseq | PRJNA290039 | Zhao et al. (2015b) | Gynophore | Developmental stages | Integrated with proteome analysis |
| RNAseq | PRJNA286040 | Peng et al. (2016) | Leaf | Two genotypes | SSR discovery |
| RNAseq | PRJNA284674 | Guimaraes et al. (2015) | Root | *A. stenosperma*/ nematode | Pest pressure time course |
| RNAseq | PRJNA251584 | Fujian Agric For Univ, unpubl. | 8 | Biotic and abiotic stress | |
| miRNAseq | PRJNA251517 | Zhao et al. (2015a) | Root | Organ | |
| RNAseq | PRJNA248910 | Chopra et al. (2014) | Leaf, pod, root | *A. hypogaea*<br>*A. ipaensis*<br>*A. duranensis* | Assembler comparisons |
| RNAseq | PRJNA243319 | Li et al. (2014) | Leaf, root, stem | Water deficit and ABA | Differential gene expression |
| RNAseq | PRJNA233534 | Geng et al. (2014) | Leaf, stem, pod, root | One genotype | Tissue-specific gene discovery |
| RNAseq | SRA053198 | Chen et al. (2013) | Pod | Above and below ground | Early embryo abortion |
| RNAseq | PRJNA185732 | Yin et al. (2013) | Seed | Two genotypes | Oil metabolism |
| RNAseq | PRJNA181974 | Xia et al. (2013), Zhong et al. (2016) | Gynophore | Developmental stages/light–dark | Differential gene expression SSR discovery |
| Small RNAseq | PRJNA146213 | Zhai et al. (2011) | Flower, nodule | Organs | Legume conservation |
| Microarray | PRJNA143823 | Chen et al. (2012) | Leaves, pods | Five varieties | Expression in response to selection |
| Microarray | PRJNA138261 | Guo et al. (2011) | Seeds | Two cultivars, *Aspergillus* infected | Differential response to *Aspergillus* exposure |
| Microarray | PRJNA123433, PRJNA123473, PRJNA119623 | Payton, unpubl. | Leaf | Two genotypes, Heat stress | Comparison of acclimated and unacclimated response |
| Microarray | PRJNA106585 | Payton et al. (2009) | Leaf, stem, peg, pod, root | Organs | Tissue-specific gene expression |
| RNAseq | PRJNA78245 | Guimaraes et al. (2012) | Leaf | *A. stenosperma*/ leaf spot | Infection time course |

(continued)

**Table 6.3** (continued)

| Data type | NCBI accession | Publication | Target tissues | Treatment/ Experimental materials[a] | Objective |
|---|---|---|---|---|---|
| RNAseq | PRJNA74747 | Guimaraes et al. (2012), Brasileiro et al. (2015) | Leaf, root | *A. duranensis*/ water deficit | Progressive water stress |
| RNAseq | PRJNA78133 | Zhang et al. (2012) | Immature seed | Three genotypes | Oil metabolism SSR discovery |
| 454 ESTs | PRJNA50587 | Nagy et al. (2012) | Roots, seeds | *A. duranensis*—two accessions | SNP discovery |
| 454 ESTs | PRJNA49471 | Guo et al. (2012) | Leaf, roots, seeds | 22 genotypes | SSR and SNP discovery |

[a]Experiments are with *A. hypogaea* unless otherwise indicated

GenBank (*A. duranensis*: GCA_000817695.1, *A. ipaensis*: GCA_000816755.1). Efforts are underway to sequence the tetraploid peanut genome, along with extensive high-density genotyping of diverse tetraploid accessions and RIL (Recombinant Inbred Lines) populations.

The primary repository for peanut genetic and genomic data for the IPGI is PeanutBase (http://peanutbase.org), which holds genome sequences, gene model sequences and annotations, genetic maps, mapped traits (QTL), expression data, germplasm information, and a variety of tools for browsing, searching, querying, and interrelating these data sets. These tools are briefly described below.

### 6.4.1 Genome Browsers

Two types of browsers are available for the diploid *Arachis* genome sequences: GBrowse (Stein et al. 2002) and JBrowse (Skinner et al. 2009). These both offer most of the same genomic data tracks mapped onto the genome sequences, but the user interface experiences and capability of these two browsers are somewhat different. Strengths of GBrowse include more intuitive views of large-scale genomic synteny with other species, richer annotations of gene models with link-outs from any gene to the gene family viewers and to close-scale synteny views with other legume species, and richer views of gene expression data for various tissues and experimental conditions. Strengths of JBrowse include a more responsive interface, with faster zooming and panning, more efficient management of tracks, and better handling of large numbers of tracks.

### 6.4.2 Sequence Search Tools

The genomic and gene model sequence can be searched via BLAST (Altschul et al. 1990, 1997) and BLAT (Kent 2002). Additionally, a keyword search function permits searching for sequence features by searching feature names and descriptions for the requested keyword.

### 6.4.3 Gene Family Viewer

The gene models for *A. duranensis* and *A. ipaensis* have been included in legume-focused gene families, which are based on the Phytozome gene families. A web module for viewing and exploring *Arachis* genes in the context of their families has been developed at LegumeInfo (http://legumeinfo.org) and is closely integrated with PeanutBase. The module provides phylogram, circular dendrogram views of the gene family phylogenetic tree.

### 6.4.4 Gene Expression Data, Atlases, and Viewers

The gene expression data sets available or linked at PeanutBase include a tissue expression atlas

**Table 6.4** Reference genes for peanut quantitative gene expression studies

| Gene | Tissues tested | Relative stability | Abundance | Publication |
|---|---|---|---|---|
| Hexameric polyubiquitin (ubq10) | Pod, seed, leaf, gynophore, root | Moderate | ++ | Brand and Hovav (2010) |
| Polyubiquitin 10 | Cotyledon, embryo axis | Moderate | +++ | Jiang et al. (2011) |
| Ubiquitin 10 | 32 tissues/stages/treatments | Moderate | +++ | Chi et al. (2012) |
| Ubiquitin | Cotyledon, embryo axis | High | +++ | Jiang et al. (2011) |
| Ubiquitin conjugating enzyme | | Moderate | + | Reddy et al. (2013) |
| Glyceraldehyde-3-phosphate dehydrogenase (gapdh) | Pod, seed, leaf, gynophore, root | Moderate | +++ | Brand and Hovav (2010) |
| glucose-6-phosphate dehydrogenase | Leaf, cotyledon, stem, root, seed/stress | High | ++ | Reddy et al. (2013) |
| RNA helicase 1 (hel1) | Pod, seed, leaf, gynophore, root | High | + | Brand and Hovav (2010) |
| Nuclear helicase | 32 tissues/stages/treatments | Low | ++ | Chi et al. (2012) |
| Yellow leaf-specific 8 gene (yls8) | Pod, seed, leaf, gynophore, root | High | ++ | Brand and Hovav (2010) |
| 14-3-3 protein (14-3-3) | Pod, seed, leaf, gynophore, root | Moderate | ++ | Brand and Hovav (2010) |
| 60 s ribosomal protein L7 (60 s) | Pod, seed, leaf, gynophore, root | High | ++ | Brand and Hovav (2010) |
| Ubiquitin C (ubc) | Pod, seed, leaf, gynophore, root | Moderate | +++ | Brand and Hovav (2010) |
| Elongation factor 1 (ef1alpha) | Pod, seed, leaf, gynophore, root | Moderate | ++ | Brand and Hovav (2010) |
| Elongation factor 1-alpha | Cotyledon, embryo axis | High | +++ | Jiang et al. (2011) |
| Elongation factor 1-beta | 32 tissues/stages/treatments | Moderate | ++ | Chi et al. (2012) |
| Elongation factor 1-beta | Leaf, cotyledon, stem, root, seed/stress | Moderate | +++ | Reddy et al. (2013) |
| Elongation factor EF-2 | Cotyledon, embryo axis | Moderate | +++ | Jiang et al. (2011) |
| Actin 7 (act7) | Pod, seed, leaf, gynophore, root | Moderate | ++ | Brand and Hovav (2010) |
| Actin | Cotyledon, embryo axis | High | + | Jiang et al. (2011) |
| Actin 11 | 32 tissues/stages/treatments | High | ++ | Chi et al. (2012) |
| Actin 11 | Leaf, cotyledon, stem, root, seed/stress | Moderate | ++ | Reddy et al. (2013) |
| Actin 2/7 | 32 tissues/stages/treatments | Low | ++ | Chi et al. (2012) |
| Alcohol dehydrogenase class III | Pod, seed, leaf, gynophore, root | High | ++ | Brand and Hovav (2010) |
| Alcohol dehydrogenase class III | Leaf, cotyledon, stem, root, seed/stress | High | ++ | Reddy et al. (2013) |
| Alpha tubulin 1 | Cotyledon, embryo axis | High | ++ | Jiang et al. (2011) |
| Alpha tubulin | 32 tissues/stages/treatments | High | ++ | Chi et al. (2012) |
| Tubulin B4 | Cotyledon, embryo axis | Low | + | Jiang et al. (2011) |
| Beta tubulin | 32 tissues/stages/treatments | Low | ++ | Chi et al. (2012) |

(continued)

**Table 6.4** (continued)

| Gene | Tissues tested | Relative stability | Abundance | Publication |
|---|---|---|---|---|
| Phospholipase D | Cotyledon, embryo axis | High | ++ | Jiang et al. (2011) |
| Malate dehydrogenase | Cotyledon, embryo axis | High | ++ | Jiang et al. (2011) |
| Ribosomal protein L24 | Cotyledon, embryo axis | nt | ++ | Jiang et al. (2011) |
| Cyclophilin | 32 tissues/stages/treatments | Moderate | ++ | Chi et al. (2012) |
| Cyclophilin | Leaf, cotyledon, stem, root, seed/stress | Moderate | ++ | Reddy et al. (2013) |
| SKP1/ask-interacting protein 16 | 32 tissues/stages/treatments | Moderate | ++ | Chi et al. (2012) |
| Metalloprotease | 32 tissues/stages/treatments | Moderate | ++ | Chi et al. (2012) |
| Phosphoenolpyruvate carboxylase-related Kinase 1 | 32 tissues/stages/treatments | Moderate | ++ | Chi et al. (2012) |
| TIP41-like family protein | 32 tissues/stages/treatments | High | ++ | Chi et al. (2012) |
| Hypothetical protein EG028875 | 32 tissues/stages/treatments | High | ++ | Chi et al. (2012) |
| Hypothetical protein GO334386 | 32 tissues/stages/treatments | High | + | Chi et al. (2012) |
| ATP synthase | Leaf, cotyledon, stem, root, seed/stress | Moderate | ++ | Reddy et al. (2013) |
| Lectin precursor | Leaf, cotyledon, stem, root, seed/stress | Low | +++ | Reddy et al. (2013) |

from Dr. Peggy Ozias-Akins and colleagues, a nematode experiment from Guimaraes et al. (2015), and a drought response experiment from Brasileiro et al. (2015).

### 6.4.5 Map Viewers

Twenty-five genetic maps for wild and cultivated peanut are available at PeanutBase at the time of writing. CMap (Youens-Clark et al. 2009) provides an interactive visual display of genetic maps and map record pages give details about each map, such as associated publications, list of linkage groups, population structure, and description of the map construction.

### 6.4.6 Quantitative Trait Locus (QTL) Search Tool

At the time of writing, 232 QTL has been curated into the PeanutBase database. Curation of peanut QTL data remains a primary focus of PeanutBase, so this number will continue to increase. QTL can be searched, examined individually via QTL record pages, and viewed on genetic maps via CMap.

### 6.4.7 Germplasm

Germplasm data available at PeanutBase includes several hundred photos of peanut germplasm, which were provided by Noelle Barkley while at the Plant Genetic Resources Conservation Center at Griffin, Georgia, USA. In addition, the GRIN database can be searched for *Arachis* germplasm from PeanutBase, and can be browsed by location using a geographic viewer.

### 6.4.8 Marker-Assisted Selection Pages

Pages developed for a few important agronomic traits integrate data, including markers and

publications into one page. The traits include Late leaf spot (*Phaeoisariopsis personata*), Peanut RKN, seed oleic acid to linoleic acid ratio, and Puccina rust.

### 6.4.9 Downloads

Full pseudomolecule sequence and the gene model genomic, CDS, and protein sequences for *A. duranensis* and *A. ipaensis* can be downloaded at PeanutBase, along with maps and tables of trait data by publication. These are all available at the download page.

## References


Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Barkley NA, Chamberlin KDC, Wang ML, Pittman RN (2011a) Genotyping and fatty acid composition analysis in segregating peanut (*Arachis hypogaea* L.) populations. Peanut Sci 38(1):11–19

Barkley NA, Wang ML, Pittman RN (2011b) A real-time PCR genotyping assay to detect FAD2A SNPs in peanuts (*Arachis hypogaea* L.). Electronic J Biotech 14(1). doi:10.2225/vol14-issue1-fulltext-12

Barkley NA, Isleib TG, Wang ML, Pittman RN (2013) Genotypic effect of ahFAD2 on fatty acid profiles in six segregating peanut (*Arachis hypogaea* L.) populations. BMC Genet 14. doi:10.1186/1471-2156-14-62

Batley J, Edwards D (2007) SNP applications in plants. In: Oraguzie NC, Rikkerink EHA, Gardiner S, deSilva HN (eds) Association mapping in plants. Springer, New York, pp 95–102

Bertioli DJ, Seijo G, Freitas FO, Valls JFM, Leal-Bertioli SCM, Moretzsohn MC (2011) An overview of peanut and its wild relatives. Plant Genet Resour-Charac Util 9:134–149

Bertioli DJ, Ozias-Akins P, Chu Y, Dantas KM, Santos SP, Gouvea E, Guimaraes PM, Leal-Bertioli SCM, Knapp SJ, Moretzsohn MC (2014) The use of SNP markers for linkage mapping in diploid and tetraploid peanuts. G3-genes genomes. Genetics 4:89–96

Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EKS, Liu X, Gao D, Clevenger J, Dash S et al (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet 48(4):46–438

Bosamia TC, Mishra GP, Thankappan R, Dobaria JR (2015) Novel and stress relevant EST derived SSR markers developed and validated in peanut. Plos One 10(6). doi:10.1371/journal.pone.0129127

Branch WD, Brenneman TB, Hookstra G (2014) Field test results versus marker assisted selection for root-knot nematode resistance in peanut. Peanut Sci 41:85–89

Brand Y, Hovav R (2010) Identification of suitable internal control genes for quantitative real-time PCR expression analyses in peanut (*Arachis hypogaea*). Peanut Sci 37:12–19

Brasileiro ACM, Morgante CV, Araujo ACG, Leal-Bertioli SCM, Silva AK, Martins ACQ, Vinson CC, Santos CMR, Bonfim O, Togawa RC, Saraiva MAP, Bertioli DJ, Guimaraes PM (2015) Transcriptome profiling of wild *Arachis* from water-limited environments uncovers drought tolerance candidate genes. Plant Mol Biol Rep 33:1876–1892

Chen Z, Wang ML, Barkley NA, Pittman RN (2010) A simple allele-specific PCR assay for detecting FAD2 alleles in both A and B genomes of the cultivated peanut for high-oleate trait selection. Plant Mol Biol Rep 28:542–548

Chen X, Hong Y, Zhang E, Liu H, Zhou G, Li S, Zhu F, Guo B, Yu J, Liang X (2012) Comparison of gene expression profiles in cultivated peanut (*Arachis hypogaea*) under strong artificial selection. Plant Breed 131:620–630

Chen X, Zhu W, Azam S, Li H, Zhu F, Li H, Hong Y, Liu H, Zhang E, Wu H, Yu S, Zhou G, Li S, Zhong N, Wen S, Li X, Knapp SJ, Ozias-Akins P, Varshney RK, Liang X (2013) Deep sequencing analysis of the transcriptomes of peanut aerial and subterranean young pods identifies candidate genes related to early embryo abortion. Plant Biotech J 11:115–127

Chen W, Jiao Y, Cheng L, Huang L, Liao B, Tang M, Ren X, Zhou X, Chen Y, Jiang H (2016) Quantitative trait locus analysis for pod- and kernel-related traits in the cultivated peanut (*Arachis hypogaea* L.). BMC Genet 17. doi:10.1186/s12863-016-0337-x

Chi X, Hu R, Yang Q, Zhang X, Pan L, Chen N, Chen M, Yang Z, Wang T, He Y, Yu S (2012) Validation of reference genes for gene expression studies in peanut by quantitative real-time RT-PCR. Mol Genet Genom 287:167–176

Chopra R, Burow G, Farmer A, Mudge J, Simpson CE, Burow MD (2014) Comparisons of de novo transcriptome assemblers in diploid and polyploid species using peanut (*Arachis* spp.) RNA-seq data. Plos One 9 (12). doi:10.1371/journal.pone.0115055

Chopra R, Burow G, Farmer A, Mudge J, Simpson CE, Wilkins TA, Baring MR, Puppala N, Chamberlin KD, Burow MD (2015) Next-generation transcriptome sequencing, SNP discovery and validation in four market classes of peanut, *Arachis hypogaea* L. Mol Genet Genom 290:1169–1180

Chu Y, Holbrook CC, Timper P, Ozias-Akins P (2007) Development of a PCR-based molecular marker to

select for nematode resistance in peanut. Crop Sci 47:841–847

Chu Y, Holbrook CC, Ozias-Akins P (2009) two alleles of ahFAD2B control the high oleic acid trait in cultivated peanut. Crop Sci 49:2029–2036

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 4:8

Chu Y, Gill R, Clevenger J, Ozias-Akins P, Timper P, Holbrook CC (2016) Identification of rare recombinants leads to tightly linked markers for nematode resistance in peanut. Peanut Sci. 43:88-93, http://dx.doi.org/10.3146/PS16-12.1

Clevenger JP, Ozias-Akins P (2015) SWEEP: a tool for filtering high-quality SNPs in polyploid crops. G3-Genes Genomes. Genetics 5:1797–1803

Clevenger J, Chavarro C, Pearl SA, Ozias-Akins P, Jackson SA (2015a) Single nucleotide polymorphism identification in polyploids: a review, example, and recommendations. Mol Plant 8:831–846

Clevenger J, Chu Y, Scheffler B, Ozias-Akins P (2016) A Developmental Transcriptome Map for Allotetraploid Arachis hypogaea. Front Plant Sci 7:1446, http://dx.doi.org/10.3389/fpls.2016.01446

Clevenger J, Chu Y, Chavarro C, Agarwal G, Bertioli DJ, Leal-Bertioli SC, Pandey MK, Vaughn J, Abernathy B, Barkley NA, Hovav R, Burow M, Nayak S, Chitikineni A, Isleib TG, Holbrook CC, Jackson SA, Varshney RK, Ozias-Akins P (2017) Genome-wide SNP Genotyping Resolves Signatures of Selection and Tetrasomic Recombination in Peanut. Mol Plant 10 (2):309-322, http://dx.doi.org/10.1016/j.molp.2016.11.015

Falleiro Suassuna TdM, Suassuna ND, Moretzsohn MdC, de Macedo Leal Bertioli SC, Bertioli DJ, de Medeiros EP (2015) Yield, market quality, and leaf spots partial resistance of interspecific peanut progenies. Crop Breed Appl Biotech 15:175–180

Faye I, Pandey MK, Hamidou F, Rathore A, Ndoye O, Vadez V, Varshney RK (2015) Identification of quantitative trait loci for yield and yield related traits in groundnut (*Arachis hypogaea* L.) under different water regimes in Niger and Senegal. Euphytica 206:631–647

Feng S, Wang X, Zhang X, Dang PM, Holbrook CC, Culbreath AK, Wu Y, Guo B (2012) Peanut (*Arachis hypogaea)* expressed sequence tag project: Progress and application. Comp Func Genom doi:10.1155/2012/373768

Fonceka D, Hodo-Abalo T, Rivallan R, Faye I, Sall MN, Ndoye O, Favero AP, Bertioli DJ, Glaszmann JC, Courtois B, Rami JF (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

Fonceka D, Tossim H-A, Rivallan R, Vignes H, Faye I, Ndoye O, Moretzsohn MC, Bertioli DJ, Glaszmann J-C, Courtois B, Rami J-F (2012a) Fostered and left behind alleles in peanut: interspecific QTL mapping reveals footprints of domestication and useful natural variation for breeding. BMC Plant Biol 12. doi:10.1186/1471-2229-12-26

Fonceka D, Tossim H-A, Rivallan R, Vignes H, Lacut E, de Bellis F, Faye I, Ndoye O, Leal-Bertioli SCM, Valls JFM, Bertioli DJ, Glaszmann J-C, Courtois B, Rami J-F (2012b) Construction of chromosome segment substitution lines in peanut (*Arachis hypogaea* L.) using a wild synthetic and QTL mapping for plant morphology. Plos One 7(11). doi:10.1371/journal.pone.0048642

Ganal MW, Altmann T, Roder MS (2009) SNP identification in crop plants. Curr Opin Plant Biol 12:211–217

Gautami B, Fonceka D, Pandey MK, Moretzsohn MC, Sujay V, Qin H, Hong Y, Faye I, Chen X, BhanuPrakash A, Shah TM, Gowda MVC, Nigam SN, Liang X, Hoisington DA, Guo B, Bertioli DJ, Rami J-F, Varshney RK (2012) An international reference consensus genetic map with 897 marker loci based on 11 mapping populations for tetraploid groundnut (*Arachis hypogaea* L.). Plos One 7(7). doi:10.1371/journal.pone.0041213

Geng L, Duan X, Liang C, Shu C, Song F, Zhang J (2014) Mining tissue-specific contigs from peanut (*Arachis hypogaea* L.) for promoter cloning by deep transcriptome sequencing. Plant Cell Physiol 55:1793–1801

Guimaraes PM, Brasileiro ACM, Morgante CV, Martins ACQ, Pappas G, Silva OB, Jr., Togawa R, Leal-Bertioli SCM, Araujo ACG, Moretzsohn MC, Bertioli DJ (2012) Global transcriptome analysis of two wild relatives of peanut under drought and fungi infection. BMC Genomics 13. doi:10.1186/1471-2164-13-387

Guimaraes PM, Guimaraes LA, Morgante CV, Silva OB, Jr., Araujo ACG, Martins ACQ, Saraiva MAP, Oliveira TN, Togawa RC, Leal-Bertioli SCM, Bertioli DJ, Brasileiro ACM (2015) Root transcriptome analysis of wild peanut reveals candidate genes for nematode resistance. Plos One 10(10). doi:10.1371/journal.pone.0140937

Guo B, Fedorova ND, Chen X, Wan C-H, Wang W, Nierman WC, Bhatnagar D, Yu J (2011) Gene expression profiling and identification of resistance genes to *Aspergillus flavus* infection in peanut through EST and microarray strategies. Toxins 3:737–753

Guo Y, Khanal S, Tang S, Bowers JE, Heesacker AF, Khalilian N, Nagy ED, Zhang D, Taylor CA, Stalker HT, Ozias-Akins P, Knapp SJ (2012) Comparative mapping in intraspecific populations uncovers a high degree of macrosynteny between A- and B-genome diploid species of peanut. BMC Genomics 13. doi:10.1186/1471-2164-13-608

Guo B, Pandey MK, He G, Zhang X, Liao B, Culbreath A, Varshney RK, Nwosu V, Wilson RF, Stalker HT (2013) Recent advances in molecular genetic linkage maps of cultivated peanut. Peanut Sci 40:95–106

Gupta SK, Baek J, Carrasquilla-Garcia N, Penmetsa RV (2015) Genome-wide polymorphism detection in peanut using next-generation restriction-site-

associated DNA (RAD) sequencing. Mol Breed 35:145

Hamidou F, Ratnakumar P, Halilou O, Mponda O, Kapewa T, Monyo E, Faye I, Ntare BR, Nigam SN, Upadhyaya HD, Vadez V (2012) Selection of intermittent drought tolerant lines across years and locations in the reference collection of groundnut (*Arachis hypogaea* L.). Field Crops Res 126:189–199

Holbrook CC, Culbreath AK (2007) Registration of 'Tifrunner' peanut. J Plant Registr 1:124

Holbrook CC, Stalker HT (2003) Peanut breeding and genetic resources. Plant Breed Rev 22:297–356

Hong Y, Pandey MK, Liu Y, Chen X, Liu H, Varshney RK, Liang X, Huang S (2015) Identification and evaluation of single-nucleotide polymorphisms in allotetraploid peanut (*Arachis hypogaea* L.) based on amplicon sequencing combined with high resolution melting (HRM) analysis. Front Plant Sci 6. doi:10.3339/fpls.2015.01068

Huang L, Jiang H, Ren X, Chen Y, Xiao Y, Zhao X, Tang M, Huang J, Upadhyaya HD, Liao B (2012) Abundant microsatellite diversity and oil content in wild *Arachis* species. Plos One 7(11). doi:10.1371/journal.pone.0050002

Huang L, He H, Chen W, Ren X, Chen Y, Zhou X, Xia Y, Wang X, Jiang X, Liao B, Jiang H (2015) Quantitative trait locus analysis of agronomic and quality-related traits in cultivated peanut (*Arachis hypogaea* L.). Theor Appl Genet 128:1103–1115

Janila P, Pandey MK, Shasidhar Y, Variath MT, Sriswathi M, Khera P, Manohar SS, Nagesh P, Vishwakarma MK, Mishra GP, Radhakrishnan T, Manivannan N, Dobariya KL, Vasanthi RP, Varshney RK (2016) Molecular breeding for introgression of fatty acid desaturase mutant alleles (ahFAD2A and ahFAD2B) enhances oil quality in high and low oil containing peanut genotypes. Plant Sci 242:203–213

Jiang S, Sun Y, Wang S (2011) Selection of reference genes in peanut seed by real-time quantitative polymerase chain reaction. Intl J Food Sci Tech 46:2191–2196

Jiang H, Huang L, Ren X, Chen Y, Zhou X, Xia Y, Huang J, Lei Y, Yan L, Wan L, Liao B (2014) Diversity characterization and association analysis of agronomic traits in a Chinese peanut (*Arachis hypogaea* L.) mini-core collection. J Integr Plant Biol 56:159–169

Jung S, Swift D, Sengoku E, Patel M, Teule F, Powell G, Moore K, Abbott AG (2000) The high oleate trait in the cultivated peanut (*Arachis hypogaea* L.) I. Isolation and characterization of two genes encoding microsomal oleoyl-PC desaturases. Mol Gen Genet 263:796–805

Katam R, Basha SM, Suravajhala P, Pechan T (2010) Analysis of peanut leaf proteome. J Proteome Res 9:2236–2254

Ke T, Yu J, Dong C, Mao H, Hua W, Liu S (2015) ocsESTdb: a database of oil crop seed EST sequences for comparative analysis and investigation of a global metabolic network and oil accumulation metabolism. BMC Plant Biol 15. doi:10.1186/s12870-014-0399-8

Kent WJ (2002) BLAT—the BLAST-like alignment tool. Genome Res 12:656–664

Khera P, Upadhyaya HD, Pandey MK, Roorkiwal M, Sriswathi M, Janila P, Guo Y, McKain MR, Nagy ED, Knapp SJ, Leebens-Mack J, Conner JA, Ozias-Akins P, Varshney RK (2013) Single nucleotide polymorphism-based genetic diversity in the reference set of peanut (*Arachis* spp.) by developing and applying cost-effective Kompetitive Allele Specific Polymerase chain reaction genotyping assays. Plant Genome 6(3). doi:10.3835/plantgenome2013.06.0019

Kottapalli KR, Rakwal R, Shibato J, Burow G, Tissue D, Burke J, Puppala N, Burow M, Payton P (2009) Physiology and proteomics of the water-deficit stress response in three contrasting peanut genotypes. Plant Cell Env 32:380–407

Leal-Bertioli SCM, Bertioli DJ, Guimaraes PM, Pereira TD, Galhardo I, Silva JP, Brasileiro ACM, Oliveira RS, Silva PIT, Vadez V, Araujo ACG (2012) The effect of tetraploidization of wild *Arachis* on leaf morphology and other drought-related traits. Env Exp Bot 84:17–24

Leal-Bertioli SCM, Cavalcante U, Gouvea EG, Ballen-Taborda C, Shirasawa K, Guimaraes PM, Jackson SA, Bertioli DJ, Moretzsohn MC (2015a) Identification of QTLs for rust resistance in the peanut wild species *Arachis magna* and the development of KASP markers for marker-assisted selection. G3-Genes Genomes. Genetics 5:1403–1413

Leal-Bertioli SCM, Moretzsohn MC, Roberts PA, Ballen-Taborda C, Borba TCO, Valdisser PA, Vianello RP, Araujo ACG, Guimaraes PM, Bertioli DJ (2015b) Genetic mapping of resistance to *Meloidogyne arenaria* in *Arachis stenosperma*: a new source of nematode resistance for peanut. G3(6):377–390

Leal-Bertioli SCM, Santos SP, Dantas KM, Inglis PW, Nielen S, Araujo ACG, Silva JP, Cavalcante U, Guimaraes PM, Brasileiro ACM, Carrasquilla-Garcia N, Penmetsa RV, Cook D, Moretzsohn MC, Bertioli DJ (2015c) *Arachis batizocoi*: a study of its relationship to cultivated peanut (*A. hypogaea*) and its potential for introgression of wild genes into the peanut crop using induced allotetraploids. Ann Bot 115:237–249

Li H-F, Zhu F-H, Li H-Y, Zhu W, Chen X-P, Hong Y-B, Liu H-Y, Wu H, Liang X-Q (2013) Proteomic identification of gravitropic response genes in peanut gynophores. J Proteomics 93:303–313

Li X, Lu J, Liu S, Liu X, Lin Y, Li L (2014) Identification of rapidly induced genes in the response of peanut (*Arachis hypogaea*) to water deficit and abscisic acid. BMC Biotech 14. doi:10.1186/1472-6750-14-58

Lopez Y, Nadaf HL, Smith OD, Connell JP, Reddy AS, Fritz AK (2000) Isolation and characterization of the delta 12-fatty acid desaturase in peanut (*Arachis hypogaea* L.) and search for polymorphisms for the high oleate trait in Spanish market-type lines. Theor Appl Genet 101:1131–1138

Macedo SE, Moretzsohn MC, M Leal-Bertioli SC, Alves DM, Gouvea EG, Azevedo VC, Bertioli DJ

(2012) Development and characterization of highly polymorphic long TC repeat microsatellite markers for genetic analysis of peanut. BMC Res Notes 5:86–86

Mondal S, Hadapad AB, Hande PA, Badigannavar AM (2014a) Identification of quantitative trait loci for bruchid (*Caryedon serratus* Olivier) resistance components in cultivated groundnut (*Arachis hypogaea* L.). Mol Breed 33:961–973

Mondal S, Hande P, Badigannavar AM (2014b) Identification of transposable element markers for a rust (*Puccinia arachidis* Speg.) resistance gene in cultivated peanut. J Phytopath 162:548–552

Mondal S, Phadke RR, Badigannavar AM (2015) Genetic variability for total phenolics, flavonoids and antioxidant activity of testaless seeds of a peanut recombinant inbred line population and identification of their controlling QTLs. Euphytica 204:311–321

Mukri G, Nadaf HL, Bhat RS, Gowda MVC, Upadhyaya HD, Sujay V (2012) Phenotypic and molecular dissection of ICRISAT mini core collection of peanut (*Arachis hypogaea* L.) for high oleic acid. Plant Breed 131:418–422

Nagy ED, Chu Y, Guo YF, Khanal S, Tang SX, Li Y, Dong WBB, Timper P, Taylor C, Ozias-Akins P, Holbrook CC, Beilinson V, Nielsen NC, Stalker HT, Knapp SJ (2010) Recombination is suppressed in an alien introgression in peanut harboring Rma, a dominant root-knot nematode resistance gene. Mol Breed 26:357–370

Nagy ED, Guo Y, Tang S, Bowers JE, Okashah RA, Taylor CA, Zhang D, Khanal S, Heesacker AF, Khalilian N, Farmer AD, Carrasquilla-Garcia N, Penmetsa RV, Cook D, Stalker HT, Nielsen N, Ozias-Akins P, Knapp SJ (2012) A high-density genetic map of *Arachis duranensis*, a diploid ancestor of cultivated peanut. BMC Genomics 13. doi:10.1186/1471-2164-13-469

Ozias-Akins P (2013) The orphan legume genome whose time has come: symposium highlights from the American Peanut Research & Education Society Annual Meeting. Peanut Sci 40:66–69

Pandey MK, Gautami B, Jayakumar T, Sriswathi M, Upadhyaya HD, Gowda MVC, Radhakrishnan T, Bertioli DJ, Knapp SJ, Cook DR, Varshney RK (2012a) Highly informative genic and genomic SSR markers to facilitate molecular breeding in cultivated groundnut (*Arachis hypogaea*). Plant Breed 131:139–147

Pandey MK, Monyo E, Ozias-Akins P, Liang X, Guimaraes P, Nigam SN, Upadhyaya HD, Janila P, Zhang X, Guo B, Cook DR, Bertioli DJ, Michelmore R, Varshney RK (2012b) Advances in *Arachis* genomics for peanut improvement. Biotech Adv 30:639–651

Pandey MK, Upadhyaya HD, Rathore A, Vadez V, Sheshshayee MS, Sriswathi M, Govil M, Kumar A, Gowda MVC, Sharma S, Hamidou F, Kumar VA, Khera P, Bhat RS, Khan AW, Singh S, Li H, Monyo E, Nadaf HL, Mukri G, Jackson SA, Guo B, Liang X, Varshney RK (2014a) Genomewide association studies for 50 agronomic traits in peanut using the 'reference set' comprising 300 genotypes from 48 countries of the semi-arid tropics of the world. Plos One 9(8). doi:10.1371/journal.pone.0105228

Pandey MK, Wang ML, Qiao L, Feng S, Khera P, Wang H, Tonnis B, Barkley NA, Wang J, Holbrook CC, Culbreath AK, Varshney RK, Guo B (2014b) Identification of QTLs associated with oil content and mapping FAD2 genes and their relative contribution to oil quality in peanut (*Arachis hypogaea* L.). BMC Genetics 15. doi:10.1186/s12863-014-0133-4

Pandey MK, Agarwal G, Kale SM, Clevenger J, Nayak SN, Sriswathi M, Chitikineni A, Chavarro C, Chen X, Upadhyaya HD, Vishwakarma MK, Leal-Bertioli S, Liang X, Bertioli DJ, Guo B, Jackson SA, Ozias-Akins P and Varshney RK. 2017. Development and evaluation of a high density genotyping 'Axiom_*Arachis*' array with 58K SNPs for accelerating genetics and breeding in groundnut. Scientific Reports, 7:40577. doi: 10.1038/srep40577

Patel M, Jung S, Moore K, Powell G, Ainsworth C, Abbott A (2004) High-oleate peanut mutants result from a MITE insertion into the FAD2 gene. Theor Appl Genet 108:1492–1502

Payton P, Kottapalli KR, Rowland D, Faircloth W, Guo BZ, Burow M, Puppala N, Gallo M (2009) Gene expression profiling in peanut using high density oligonucleotide microarrays. BMC Genomics 10

Peng Z, Gallo M, Tillman BL, Rowland D, Wang J (2016) Molecular marker development from transcript sequences and germplasm evaluation for cultivated peanut (*Arachis hypogaea* L.). Mol Genet Genom 291:363–381

Poland JA, Rife TW (2012) Genotyping-by-Sequencing for plant breeding and genetics. Plant Genome 5:92–102

Qin H, Feng S, Chen C, Guo Y, Knapp S, Culbreath A, He G, Wang ML, Zhang X, Holbrook CC, Ozias-Akins P, Guo B (2012) An integrated genetic linkage map of cultivated peanut (*Arachis hypogaea* L.) constructed from two RIL populations. Theor Appl Genet 124:653–664

Rafalski JA (2002) Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Sci 162:329–333

Ravi K, Vadez V, Isobe S, Mir RR, Guo Y, Nigam SN, Gowda MVC, Radhakrishnan T, Bertioli DJ, Knapp SJ, Varshney RK (2011) Identification of several small main-effect QTLs and a large number of epistatic QTLs for drought tolerance related traits in groundnut (*Arachis hypogaea* L.). Theor Appl Genet 122:1119–1132

Reddy DS, Bhatnagar-Mathur P, Cindhuri KS, Sharma KK (2013) Evaluation and validation of reference genes for normalization of quantitative real-time PCR based gene expression studies in peanut. Plos One 8(10). doi:10.1371/journal.pone.0078555

Robledo G, Seijo G (2010) Species relationships among the wild B genome of *Arachis* species (section *Arachis*) based on FISH mapping of rDNA loci and heterochromatin detection: a new proposal for genome arrangement. Theor Appl Genet 121:1033–1046

Samoluk SS, Chalup L, Robledo G, Seijo JG (2015) Genome sizes in diploid and allopolyploid *Arachis* L. species (section *Arachis*). Genet Res Crop Evol 62:747–763

Sharma S, Upadhyaya HD, Varshney RK, Gowda CLL (2013) Pre-breeding for diversification of primary gene pool and genetic enhancement of grain legumes. Front Plant Sci 4. doi:10.3389/fpls.2013.00309

Shilman F, Brand Y, Brand A, Hedvat I, Hovav R (2011) Identification and molecular characterization of homeologous delta 9-stearoyl acyl carrier protein desaturase 3 genes from the allotetraploid peanut (*Arachis hypogaea*). Plant Mol Biol Rep 29:232–241

Shirasawa K, Hirakawa H, Tabata S, Hasegawa M, Kiyoshima H, Suzuki S, Sasamoto S, Watanabe A, Fujishiro T, Isobe S (2012a) Characterization of active miniature inverted-repeat transposable elements in the peanut genome. Theor Appl Genet 124:1429–1438

Shirasawa K, Koilkonda P, Aoki K, Hirakawa H, Tabata S, Watanabe M, Hasegawa M, Kiyoshima H, Suzuki S, Kuwata C, Naito Y, Kuboyama T, Nakaya A, Sasamoto S, Watanabe A, Kato M, Kawashima K, Kishida Y, Kohara M, Kurabayashi A, Takahashi C, Tsuruoka H, Wada T, Isobe S (2012b) In silico polymorphism analysis for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. BMC Plant Biol 12. doi:10.1186/1471-2229-12-80

Shirasawa K, Bertioli DJ, Varshney RK, Moretzsohn MC, Leal-Bertioli SCM, Thudi M, Pandey MK, Rami J-F, Fonceka D, Gowda MVC, Qin H, Guo B, Hong Y, Liang X, Hirakawa H, Tabata S, Isobe S (2013) integrated consensus map of cultivated peanut and wild relatives reveals structures of the A and B Genomes of *Arachis* and divergence of the legume genomes. DNA Res 20:173–184

Shirasawa K, Isobe S, Tabata S, Hirakawa H (2014) Kazusa Marker database: a database for genomics, genetics, and molecular breeding in plants. Breed Sci 64:264–271

Simpson CE (2001) Use of wild *Arachis* species/introgression of genes into *A. hypogaea* L. Peanut Sci 28:114–116

Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH (2009) JBrowse: a next-generation genome browser. Genome Res 19:1630–1638

Stalker HT, Tallury SP, Ozias-Akins P, Bertioli D, Bertioli SCL (2013) The value of diploid peanut relatives for breeding and genomics. Peanut Sci 40:70–88

Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, Lewis S (2002) The generic genome browser: a building block for a model organism system database. Genome Res 12:1599–1610

Sun Y, Wang Q, Li Z, Hou L, Dai S, Liu W (2013) Comparative proteomics of peanut gynophore development under dark and mechanical stimulation. J Proteome Res 12:5502–5511

Temsch EM, Greilhuber J (2000) Genome size variation in *Arachis hypogaea* and *A. monticola* re-evaluated. Genome 43:449–451

Varshney RK (2016) Exciting journey of 10 years from genomes to fields and markets: some success stories of genomics-assisted breeding in chickpea, pigeonpea and groundnut. Plant Sci 242:98–107

Varshney RK, Pandey MK, Janila P, Nigam SN, Sudini H, Gowda MVC, Sriswathi M, Radhakrishnan T, Manohar SS, Nagesh P (2014) Marker-assisted introgression of a QTL region to improve rust resistance in three elite and popular varieties of peanut (*Arachis hypogaea* L.). Theor Appl Genet 127:1771–1781

Wang ML, Sukumaran S, Barkley NA, Chen Z, Chen CY, Guo B, Pittman RN, Stalker HT, Holbrook CC, Pederson GA, Yu J (2011) Population structure and marker-trait association analysis of the US peanut (*Arachis hypogaea* L.) mini-core collection. Theor Appl Genetics 123:1307–1317

Wang H, Penmetsa RV, Yuan M, Gong L, Zhao Y, Guo B, Farmer AD, Rosen BD, Gao J, Isobe S, Bertioli DJ, Varshney RK, Cook DR, He G (2012) Development and characterization of BAC-end sequence derived SSRs, and their incorporation into a new higher density genetic map for cultivated peanut (*Arachis hypogaea* L.). BMC Plant Biol 12. doi:10.1186/1471-2229-12-10

Wang H, Pandey MK, Qiao L, Qin H, Culbreath AK, He G, Varshney RK, Scully BT, Guo B (2013a) Genetic mapping and quantitative trait loci analysis for disease resistance using F-2 and F-5 generation-based genetic maps derived from 'Tifrunner' x 'GT-C20' in peanut. Plant Genome 6(3). doi:10.3835/plantgenome2013.05.0018

Wang ML, Chen CY, Tonnis B, Barkley NA, Pinnow DL, Pittman RN, Davis J, Holbrook CC, Stalker HT, Pederson GA (2013b) Oil, fatty acid, flavonoid, and resveratrol content variability and FAD2A functional SNP genotypes in the US peanut mini-core collection. J Agric Food Chem 61:2875–2882

Wang T, Chen X-P, Li H-F, Liu H-Y, Hong Y-B, Yang Q-L, Chi X-Y, Yang Z, Yu S-L, Li L, Liang X-Q (2013c) Transcriptome identification of the resistance-associated genes (RAGs) to *Aspergillus flavus* infection in pre-harvested peanut (*Arachis hypogaea*). Func Plant Biol 40:292–303

Wang Y, Zhang X, Zhao Y, Prakash CS, He G, Yin D (2015a) Insights into the novel members of the FAD2 gene family involved in high-oleate fluxes in peanut. Genome 58:375–383

Wang ML, Tonnis B, An Y-QC, Pinnow D, Tishchenko V, Pederson GA (2015b) Newly identified natural high-oleate mutant from *Arachis hypogaea* L.

subsp *hypogaea*. Mol Breed 35(9). doi:10.1007/s11032-015-0377-3

Wang ML, Khera P, Pandey MK, Wang H, Qiao L, Feng S, Tonnis B, Barkley NA, Pinnow D, Holbrook CC, Culbreath AK, Varshney RK, Guo B (2015c) Genetic mapping of QTLs controlling fatty acids provided insights into the genetic control of fatty acid synthesis pathway in peanut (*Arachis hypogaea* L.). Plos One 10(4). doi:10.1371/journal.pone.0119454

Xia H, Jiang N, Hou L, Zhang Y, Li C, Li A, Zhao C (2015) Data in support of proteome analysis of gynophores and early swelling pods of peanut (*Arachis hypogaea* L.). Data in Brief 5:1056–1059. doi:10.1016/j.dib.2015.11.029

Yeri SB, Shirasawa K, Pandey MK, Gowda MVC, Sujay V, Shriswathi M, Nadaf HL, Motagi BN, Lingaraju S, Bhat ARS, Varshney RK, Krishnaraj PU, Bhat RS (2014) Development of NILs from heterogeneous inbred fa resistance QTL in peanut (*Arachis hypogaea* L.). Plant Breed 133:80–85

Yin D, Wang Y, Zhang X, Li H, Lu X, Zhang J, Zhang W, Chen S (2013) De novo assembly of the peanut (*Arachis hypogaea* L.) seed transcriptome revealed candidate unigenes for oil accumulation pathways. Plos One 8(9). doi:10.1371/journal.pone.0073767

Youens-Clark K, Faga B, Yap IV, Stein L, Ware D (2009) CMap 1.01: a comparative mapping application for the Internet. Bioinformatics 25:3040–3042

Zhai J, Jeong D-H, De Paoli E, Park S, Rosen BD, Li Y, Gonzalez AJ, Yan Z, Kitto SL, Grusak MA, Jackson SA, Stacey G, Cook DR, Green PJ, Sherrier DJ, Meyers BC (2011) MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. Genes Devel 25:2540–2553

Zhang J, Liang S, Duan J, Wang J, Chen S, Cheng Z, Zhang Q, Liang X, Li Y (2012) De novo assembly and characterisation of the transcriptome during seed development, and generation of genic-SSR markers in peanut (*Arachis hypogaea* L.). BMC Genomics 13. doi:10.1186/1471-2164-13-90

Zhao Y, Prakash CS, He G (2012) Characterization and compilation of polymorphic simple sequence repeat (SSR) markers of peanut from public database. BMC Res Notes 5:362. doi:10.1186/1756-0500-5-362

Zhao C, Xia H, Cao T, Yang Y, Zhao S, Hou L, Zhang Y, Li C, Zhang X, Wang X (2015a) Small RNA and degradome deep sequencing reveals peanut microRNA roles in response to pathogen infection. Plant Mol Biol Rep 33:1013–1029

Zhao C, Zhao S, Hou L, Xia H, Wang J, Li C, Li A, Li T, Zhang X, Wang X (2015b) Proteomics analysis reveals differentially activated pathways that operate in peanut gynophores at different developmental stages. BMC Plant Biol 15. doi:10.1186/s12870-015-0582-6

Zhong R, Zhou M, Zhao C, Hou L, Li C, Wang X, Tang R, Xia H (2016) SSR marker development from peanut gynophore transcriptome sequencing. Plant Breed 135:111–117

Zhou X, Xia Y, Ren X, Chen Y, Huang L, Huang S, Liao B, Lei Y, Yan L, Jiang H (2014) Construction of a SNP-based genetic linkage map in cultivated peanut based on large scale marker development using next-generation double-digest restriction-site-associated DNA sequencing (ddRADseq). BMC Genomics 15. doi:10.1186/1471-2164-15-351

# Classical and Molecular Approaches for Mapping of Genes and Quantitative Trait Loci in Peanut

7

Manish K. Vishwakarma, Spurthi N. Nayak, Baozhu Guo, Liyun Wan, Boshou Liao, Rajeev K. Varshney and Manish K. Pandey

**Abstract**

Advances in availability of genomic resources coupled with genetic resources have accelerated the process of developing better understanding of cytogenetics and genetics of peanut using modern technologies. The cytogenetic studies provided greater insights on chromosomal structures and behaviour of different *Arachis* species along with their genetic relationship with each other. Researchers are moving faster now in using single nucleotide polymorphism (SNP) markers in their genetic studies as simple sequence repeats (SSRs) did not provide optimum genome density for genetic mapping studies in peanut. Due to availability of reference genome of diploid progenitors, resequencing of some genotypes and soon to be available tetraploid genome, a high-density genotyping array with 58 K SNPs is now available for conducting high-resolution mapping in peanut. ICRISAT has developed next generation genetic mapping populations such as multi-parent advanced generation intercross (MAGIC) and nested association mapping (NAM) populations for conducting high-resolution trait mapping for multiple traits in one go. Affordability of sequencing also encouraged initiation of sequence-based trait mapping such as QTL-seq for dissecting foliar disease resistance trait. Few successful examples are available in peanut regarding development of diagnostic markers and their deployment in breeding to develop improved genotypes, which may see a significant increase in coming years for developing appropriate genomics tools for breeding in peanut.

M.K. Vishwakarma · S.N. Nayak · R.K. Varshney · M.K. Pandey (✉)
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India
e-mail: m.pandey@cgiar.org

B. Guo
USDA-ARS, Crop Protection and Management Research Unit, Tifton, USA

L. Wan
Oil Crops Research Institute (OCRI) of the Chinese Academy of Agricultural Sciences, Wuhan, China

B. Liao
Crops Research Institute of the Chinese Academy of Agricultural Sciences, Wuhan, China


R.K. Varshney et al. (eds.), *The Peanut Genome*, Compendium of Plant Genomes, DOI 10.1007/978-3-319-63935-2_7

## 7.1 Introduction

Cultivated peanut (*Arachis hypogaea* L.), also called as groundnut, is the second largest oilseed legume crop after soybean in the world. It covers the tropical and subtropical regions but primarily grown in the semi-arid tropics (SAT) regions of the world. This crop is cultivated in >100 countries of Africa, Asia and Americas and is consumed in almost all the countries in one or the other form. The global annual production in 2014 was 42.31 metric tons from an area of 25.44 million hectares (FAO 2014, accessed on 10th March 2016).

Peanut productivity is highly affected by several biotic and abiotic stresses across the world. The major abiotic stresses include terminal drought, heat and salinity (Fig. 7.1). The major biotic stresses include rust, early leaf spot (ELS) and late leaf spot (LLS), tomato spotted wilt virus (TSWV), groundnut rosette disease (GRD), peanut clump virus disease (PCVD), peanut stripe virus (PStV), peanut bud necrosis disease (PBND), peanut stem necrosis disease (PSND), bacterial wilt and root-knot nematodes (Nigam et al. 2012). The above-mentioned stresses cause massive yield loss in different intensities and quantity in addition to deteriorating the quality of the produce. In addition to above yield reducing stresses, *Aspergillus flavus* infection is a very serious issue as it produces carcinogenic mycotoxins known as aflatoxins which have an adverse impact on human health and the economy. Aflatoxin is known to cause cancer in human beings, animals and poultry birds that are fed with contaminated peanut seeds/cakes. The major causal agent for aflatoxin contamination, *A. flavus,* is predominant species in Asia and Africa while *A. parasiticus* in the USA (see Pandey et al. 2012a; Janila and Nigam 2013). Further, recent increased awareness among consumers has also raised demand for good oil quality and nutritious peanut seed.

The domesticated peanut is an allotetraploid ($2n = 4x = 40$) crop with two subgenomes

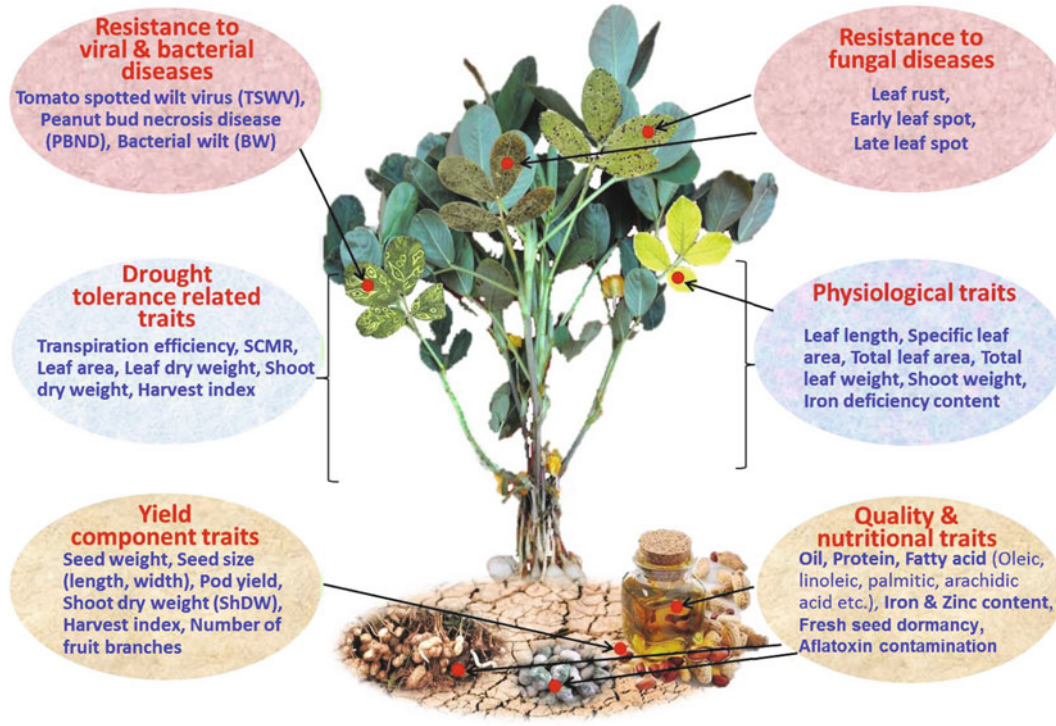


**Fig. 7.1** Trait mapping efforts in peanut for developing trait-linked markers

(A and B). About 3500 years ago, these two subgenomes believed to have come together from a single hybridization event between two diploid wild species ($2n = 2x = 20$) accompanied by whole genome duplication. The above event gave rise to cultivated peanut which then remained isolated over the centuries from its progenitors' genepool for further diversification. As a consequence, the limited evolutionary history, coupled with hybridization barriers between diploids and the tetraploid, have created roadblocks in the mobilizing alleles from wild species to the cultivated peanuts leading to narrow genetic base for today's cultivated peanut (Simpson 1991). Amalgamation of the genomics with the integrated classical breeding has calibre to boost the yield of peanut by overcoming selected genetic barriers. Since a decennary, enormous progress has been made in the peanut genomics leading to the development of enormous genetic and genomic resources such as genome sequences, whole genome re-sequencing (WGRS), molecular markers, mapping populations, genetic maps, high throughput sequencing and genotyping platforms, transcriptome sequencing and proteome (Pandey et al. 2012a, 2016a; Varshney et al. 2013, 2015a, b). These resources have been exploited and utilized in genetic map construction, quantitative trait loci (QTL) mapping for traits, association mapping and ultimately transform it in the translational genomics for the improvement of peanut (Pandey et al. 2014a, 2016a; Varshney et al. 2015b).

Genetics and genomics offer excellent opportunity to accelerate genetic gains and achieve developing improved peanut varieties with high yield and quality. In the case of peanut, such technologies will contribute to improving the biotic and abiotic resistance, oil quality, seed quality, seed nutrition and yield. For increasing genetic gains in breeding programs, efficient utilization of genetic resources conserved in available germplasm through genomics approaches is essential. Development of superior varieties with the improved characteristics keeping in mind the requirements of a specific environment, growers and consumers will eventually enhance the chances of adoption, which unfortunately touches quiet low now. Genomics utilizes the analysis of full genetic constitute by tagging, sequencing and functional examination to discover genes/QTLs that operate, check and alter the expression. Plant breeding along with genetics and genomics is a potent way to give phenomenal growth to agriculture productivity and sustainability. Advances in next-generation sequencing (NGS) technologies has accelerated the pace in crop genetics and breeding (Varshney et al. 2009a). Peanut Genome Consortium (PGC) with the collaboration of international partners initiated the International Peanut Genome Initiative (IPGI) in 2012 and released the first chromosomal-scale draft sequences of two progenitors of tetraploid cultivated peanut (*A. hypogaea*), representing A-genome (*Arachis duranensis,* accession V14167) and B-genome (*A. ipaensis,* accession K30076) (Bertioli et al. 2016). In a parallel effort by Diploid Progenitor A-genome Sequencing Consortium (DPPAGSC) (http://ceg.icrisat.org/dppga/Manuscript.html), another draft sequence of A-genome progenitor (*A. duranensis*, accession PI475845) has also been developed and made available in the public domain (Chen et al. 2016a). The IPGI-led genome assembly of A-genome progenitor is better than the DPPAGSC-led genome assembly in terms of quality and applicability in further peanut genomics research. Nevertheless, the DPPAGSC-led genome assembly provided in-depth genome analysis identifying genes for geocarpy, oil biosynthesis and allergens. The above-mentioned genome assemblies will further enhance the genomics research leading to gene discovery, high-resolution trait mapping and molecular breeding.

This chapter provides updates on cytological studies, molecular markers, genetic linkage maps and trait linked QTL identification using linkage and association mapping/linkage disequilibrium mapping approaches. Also, we discuss the development of complex high-resolution trait mapping populations like MAGIC (multi-parent advanced generation intercross) and NAM (nested association mapping). In addition, we

discussed NGS-based SNPs identification linked to gene/QTLs for concerned traits using modern high-resolution trait mapping and gene discovery approaches.

## 7.2 Advances in Peanut Cytological Research

With the generation of huge data through NGS technologies, the challenge comes in computational analysis. The advanced plant cytogenetics has made essential contributions to genomics by interpreting the scaffolds, marker orders, genome arrangements like translocations and inversions. Chromosome markers developed using fluorescent in situ hybridization (FISH) with rDNA probes and fluorescent banding were used in development of chromosome map of peanut (Robledo and Seijo 2010). FISH is commonly used to map unique or low copy number sequences and to localize repetitive sequence to produce chromosome recognition cocktails or explore genome relations in polyploid or closely related plant species. Chromosome identification in peanuts started with studies carried out by Husted (1933), who delineated the occurrence of two pairs of chromosomes in peanut. Karyotyping analysis and relationships among varieties of *A. hypogaea* L. were studied (Stalker and Dalmacio 1986) and later, the relationship of *Arachis* section was cytologically implicated (Stalker et al. 1991). Development of fluorescent banding patterns (like Q-, C-, G-, R-, T-banding) revolutionized the karyotyping and characterization of the genomes of different plant species. The fluorochrome banding patterns acted like markers to differentiate different species of *Arachis* section (Raina and Mukai 1999; Seijo et al. 2004). Modified genomic in situ hybridization (GISH) techniques were used to study the genomic relationships between the cultivated peanut and its probable progenitors (Seijo et al. 2007). Lately, sequential GISH-FISH method was utilized to study the chromosome analysis of peanut (Pei et al. 2015).

Cytogenetic studies have been very important to distinguish and define different genomes of *Arachis* section including the first genome constitution establishment within the *Arachis* genus. Based on thorough cytological studies, the *Arachis* species were categorized to have A genome and non-A (B, D, E, F, K, P) genomes (Smartt et al. 1978; Smartt and Stalker 1982; Stalker 1991; Robledo and Seijo 2010). Interestingly, the species within each subgroup were more closely distributed geographically and were named using geographical reference (Robledo et al. 2009). The karyotype features of *A. duranensis* and *A. cardenasii* indicated the occurrence of 'A' genome; *A. ipaensis* of 'B' genome; *A. glandulifera* of 'D' genome; *A. batizocoi*, *A. cruziana*, *A. krapovickasii* of 'K' genome; and *A. benensis*, *A. trinitensis* of 'F' genome. The origin and evolution of peanuts have been studied based on the rDNA, internal transcribed spacer (ITS) region, restriction fragment length polymorphism (RFLP) markers. *A. duranensis* and *A. ipaensis* were proposed to be the probable progenitors of cultivated peanut using RFLP analysis (Burow et al. 2009). Later based on the studies of rDNA using FISH and heterochromatin distribution showed that the A genome of the cultivated peanut was more related to *A. duranensis* and B genome was related to *A. ipaensis* (Seijo et al. 2004; Robledo and Seijo 2010).

With advances in technologies and modification of existing technologies like spectral karyotyping (SKY) and DNA fibre-FISH can be used in more accurate physical mapping. During SKY, all the chromosomes can be simultaneously visualized using chromosome specific probes (Imataka and Arisaka 2012). In DNA fibre-FISH technique, the extended DNA fibres released from lyzed nuclei are used as specimens for hybridization instead of nuclei or chromosomes as for conventional FISH. Fibre FISH can distinguish two probes separated by 1 kb on a DNA fibre (Wang et al. 2013). Although there are no reports yet in peanut, there is high scope to use these new techniques in characterizing the wild relatives. These crop wild relatives (CWR) have untapped genomic regions that confer resistance

to biotic stresses like ELS, LLS, PBND, PStV, PMV, TSWV, aflatoxin, corn ear worm, southern corn root worm, thrips, leaf hoppers and Spodoptera (Dwivedi and Johri 2003, 2008; Upadhyaya et al. 2011).

QTLs for disease resistance were reported from crosses involving wild diploid species in peanut (Bertioli et al. 2009) and the derivatives from the wild (Gowda et al. 2002). Besides, introgression of disease resistance genes has also been reported from the wild diploid species (*A. cardenasii*) into an elite peanut variety (Simpson and Starr 2003; Simpson et al. 2003). Very recently, introgression of rust resistance region into three elite cultivars was reported, where the donor GPBD 4 is the second cycle derivative of interspecific hybridization (Varshney et al. 2014). The introgression lines can be used for breeding and mapping of genes/QTLs simultaneously (Alpert and Tanksley 1996; Tanksley et al. 1996; Tanksley and McCouch 1997) through advanced backcross-QTL (AB-QTL) approach (Iyer-Pascuzzi et al. 2007). Synthetics developed from wild species were used to develop chromosome segment substitution line (CSSL) and used to dissect plant morphology in case of peanut (Fonceka et al. 2012). Further, CSSL intercosses and CS-RILs can be used to dissect the complex traits involved in resistance to biotic and abiotic stresses.

## 7.3 Genetic Markers

DNA markers have played very crucial role in forming backbone of genomics, with the utilization in genetic mapping, genomic assisted breeding (GAB), association studies, genomic selection and fine mapping (Hyten et al. 2010). During the 90s, apart from DNA marker, isozyme a protein-based marker system was deployed for the genetic diversity in peanut (Lacks and Stalker 1993). Shortly, DNA-based marker systems viz. RFLP (Kochert et al. 1996), random amplified polymorphic DNA (RAPD) (Subramanian et al. 2000) and amplified fragment length polymorphism (AFLP) (He and Prakash 1997; Herselman 2003) replaced isozyme completely. Meanwhile, several other DNA-based markers also reported like sequence-related amplified polymorphism (SRAP) (Wang et al. 2010), single strand conformational polymorphism (SSCP) (Nagy et al. 2010), and miniature inverted-repeat transposable elements (MITEs) (Shirasawa et al. 2012a). Notwithstanding, they were rarely utilized for peanut genotyping. With the most promising and reliable technology, attention of peanut researchers shifted towards development of simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs) which rely on sequence information.

SSR markers are regarded as the marker of choice, because of several worthy properties viz. co-dominance, reproducibility, high variability, broad genome coverage and easy to use (Gupta and Varshney 2000). Development of SSR markers in *Arachis* came into existence in the year 1999, although in very less number, i.e. 26 SSRs yet an important initiative in peanut genomics studies (Hopkins et al. 1999). Nevertheless, >15,000 genomic as well as genic SSRs have been developed in peanut in last 15 years (Guo et al. 2013; Shirasawa et al. 2012a; Pandey et al. 2016a). Several of these markers are still not available to the global peanut research community. Few studies were also carried out to check usefulness of these markers by checking polymorphism in different germplasm sets including parents of mapping population, construction of genetic maps, marker-trait association analysis and also molecular breeding (Pandey et al. 2012b; Varshney et al. 2013).

A total of 199 highly informative SSRs with >0.50 PIC were reported after screening 4485 SSR markers (Pandey et al. 2012b). This study also reported >900 novel SSR markers, which were made accessible to the global peanut research community. Similarly, Zhao et al. (2012) and Macedo et al. (2012) reported 1343 and 78 polymorphic SSRs after screening 9274 and 146 SSRs, respectively. Recently, by using EST database available in public domain, Peng et al. (2016) reported development of 6455 SSR markers, of which only 339 SSRs amplified and only 22 were found polymorphic. Thanks to the genome

sequencing effort which has now ended the scarcity of genetic markers in peanut and now lakhs of SSRs are available due availability of draft genome sequences of the ancestor genomes (Bertioli et al. 2016; Chen et al. 2016a). Two research groups exploited the reference genome of diploid progenitors and made available primers for >150K SSRs (Luo et al. 2017; Zhao et al. 2017).

SNPs acts as a direct marker because a nucleotide base is the smallest unit of inheritance, the sequence information provides the accurate nature of the allelic variation. This sequence variation affects the development of the organism and their response to the environment. Ample amount of SNPs are dispersed in the genome, one SNPs at each 100–300 bp (Gupta et al. 2001). SNP markers are invaluable as a tool for genome wide association analysis and genomic selection offering the potential for generating ultra-high-density genetic maps. SNP development is difficult in peanut due to allotetraploidy that provides the abundance of polymorphism amongst homoeologous genomes, i.e. A and B genome (Dwivedi et al. 2007). Accordingly, development of SNP is very low in peanut. A 1536 SNPs Illumina GoldenGate array were designed by comparing sequences of 17 tetraploid genotypes with Tifrunner's transcriptome in the University of Georgia (Nagy et al. 2012).

In order to deploy SNPs in breeding program, kompetitive allele-specific polymerase chain reaction (KASP) assay markers appear as a good choice and cost effective. In this context, KASP assay for 90 SNPs were developed and deployed for genetic diversity in a very diverse germplasm panel, i.e. 'Reference Set' (Khera et al. 2013). Likewise, at the University of California-Davis, 768-SNP Illumina GoldenGate array was developed (see Varshney et al. 2013). Also, robustness of KASP assays was validated for SNPs in peanut (Chopra et al. 2015). A high-resolution melting, also used for an alternative SNP assay to validate SNPs in peanut (Hong et al. 2015). Nonetheless, Genotyping-by-sequencing (GBS) is another approach based on NGS technology which led one step forward to mine SNP markers for use in genetic analyses and genotyping (Beissinger et al. 2013). It is a low cost technology where there is less sample handling, PCR and purification steps and multiplexed based on precise barcoding (Davey et al. 2011). In peanut, GBS has been well demonstrated where the SNP markers developed were used to construct linkage map and QTL analysis in cultivated peanut (Zhou et al. 2014; Zhao et al. 2016). Completion of genome sequencing in 2016 for both the diploid progenitors will now facilitate SNP discovery in large scale. Recently, ICRISAT together with University of Georgia has developed SNP array with 58,000 informative SNPs (Pandey et al. 2017a). Development of such array will now facilitate generation of high-density genotyping data and high-resolution genetic mapping for trait discovery and diagnostic marker development for trait of interest. Such high throughput genotyping system will also facilitate deployment of modern breeding approaches in peanut wherein genome-wide SNP-based genotyping is essential for achieving higher genetic gains with more precision.

## 7.4 Genetic Linkage Maps for Diploid and Tetraploid Peanuts

Identification of molecular markers leads to construction of genetic maps and detection of genes/QTLs. Since last decennary, numbers of mapping populations and linkage maps have been developed for diploids (A and B sub-genome) and tetraploid (AABB-genome) peanut (Pandey et al. 2012a). Notwithstanding, these genetic maps were not up to mark as they had less number of markers and low density. Now, NGS-based techniques are available to identifying SNPs to enrich these maps with more number of markers and density (Table 7.1).

### 7.4.1 Diploid Genetic Maps

Genetic map construction was first initiated for the diploid (AA) genome leading to development of five separate maps using $F_2$ population by deploying a range of markers such as RFLP, AFLP and RAPD, and later SNP markers in peanut

**Table 7.1** List of genetic maps constructed in the *Arachis* species using different molecular markers

| S. N. | Genome | Population used | Marker loci mapped | Marker type | Linkage groups | Total map distance (cM) | Inter-marker distance | References |
|---|---|---|---|---|---|---|---|---|
| 1 | AA | F2 | 102–1724 | AFLP, RFLP, SNP, SSR, SCAR, SSCP | 10–12 | 1063–1230.89 | 3.88–10.47 | Halward et al. (1993), Milla (2003), Moretzsohn et al. (2005), Nagy et al. (2012) |
| | | BC1F1 | 384–437 | RAPD, RFLP | 11 | 800 | 3.88 | Garcia et al. (2005) |
| | | F5 | 170–1724 | AFLP, RFLP, SSR, SCAR, SNP, SSCP, transposon | 10–11 | 544–705.10 | 1.24–1.84 | Shirasawa et al. (2013), Bertioli et al. (2014) |
| 2 | BB | F2 | 149–449 | SSR | 10–16 | 1278.6–1294 | 2.84–8.68 | Moretzsohn et al. (2009), Guo et al. (2012) |
| | | F6 | 680 | SSR, transposon | 10 | 461 | 0.68 | Shirasawa et al. (2013) |
| 3 | AABB | F2 | 12–1452 | AFLP, SRAP, SSR, DArT, DArTseq, transposon | 5–22 | 139.40–3525.8 | 1.7–11.61 | Herselman et al. (2004), Wang et al. (2012, 2013), Shirasawa et al. (2012a, b), Vishwakarma et al. (2016), Chen et al. (2016b), Shasidhar et al. (2017) |
| | | BC1F1 | 298 and 370 | RFLP, SSR | 21–23 | 1843.7–2210.0 | 5.97–6.18 | Burow et al. (2001), Foncéka et al. (2009) |
| | | RILs | 29–1685 | SSR, CAPS, SNP | 8–26 | 401.7–2208.2 | 0.85–18.55 | Jiang et al. (2007), Hong et al. (2008, 2009), Varshney et al. (2009b), Peng et al. (2010), Ravi et al. (2011), Sarvamangala et al. (2011), Hong et al. (2010), Khedikar et al. (2010), Sujay et al. (2012), Gautami et al. (2012a, b), Qin et al. (2012), Mondal et al. (2012), Zhou et al. (2014), Mondal et al. (2014a, b) |

(continued)

**Table 7.1** (continued)

| S. N. | Genome | Population used | Marker loci mapped | Marker type | Linkage groups | Total map distance (cM) | Inter-marker distance | References |
|---|---|---|---|---|---|---|---|---|
| | | F6 | 772–1261 | SNP, SSR and transposon | 20 | 1442–1487.3 | 1.14–1.92 | Shirasawa et al. (2013), Bertioli et al. (2014) |
| | | F8 | 237 | SNP, SSR | 20 | 1627.4 | 6.8 | Zhao et al. (2016) |
| 4 | **Integrated genetic maps** | | | | | | | |
| | AABB | 3 RILs | 175 | – | 22 | 885.4 | 5.06 | Hong et al. (2010) |
| | AABB | 2 RILs | 225 | – | 20 | 1152.9 | 5.12 | Sujay et al. (2012) |
| | AABB | 2 RILs | 293 | – | 20 | 2840.8 | 9.70 | Gautami et al. (2012a) |
| | AABB | 2 RILs | 324 | – | 21 | 1352.1 | 4.17 | Qin et al. (2012) |
| | AABB | 10 RILs and 1 BC | 897 | – | 20 | 3863.6 | 4.31 | Gautami et al. (2012b) |
| | AA, BB & AABB | 3 RILs | 3693 | – | 20 | 2651 | 0.72 | Shirasawa et al. (2013) |

(Halward et al. 1993; Milla 2003; Moretzsohn et al. 2005; Nagy et al. 2012). These maps had 10–12 linkage groups (LGs) with map distance ranging from 1063 to 1231 cM, and 3.88 to 10.47 cM inter-marker distance. Subsequently, by using $BC_1F_1$ population for the diploid (AA genome), a genetic map was developed with 206 RFLP and RAPD marker loci spanning 800 cM distance and 3.88 cM inter-marker distance (Garcia et al. 2005). Later, two more genetic maps were constructed using the $F_5$ generation with the AFLP, RFLP, SSR, SCAR, SNP and SSCP markers and achieved 544–705.10 cM map distance and 1.24 to 1.84 cM inter-marker distance (Shirasawa et al. 2013; Bertioli et al. 2014). These genetic maps played an important role in several genetic studies including development of peanut A-genome assembly.

For the diploid (BB genome), initially $F_2$ population was used to construct SSR-based genetic map with 10–16 LGs which covered the map distance ranging from 1278.6 to 1294 cM and inter-marker distances 2.84 to 8.68 cM, respectively (Moretzsohn et al. 2009; Guo et al. 2012). Later in $F_6$ generation map was constructed with the 461 SSR and transposon markers and inter-marker distance was reduced up to 0.68 cM (Shirasawa et al. 2013).

### 7.4.2 Tetraploid Genetic Maps

Most of the linkage map construction work was done for the tetraploid genome by considering various mapping populations such as $F_2$, $BC_1F_1$ and recombinant inbred lines (RILs). In the $F_2$ population, several marker systems were deployed to construct maps viz. AFLP, SRAP, SSR, DArT, DArTseq and transposon, and achieved map distance up to 3526 cM with the inter-marker distance of 1.7–11.6 cM (Herselman et al. 2004; Wang et al. 2012, 2013; Shirasawa et al. 2012b; Chen et al. 2016b, Vishwakarma et al. 2016). Using $BC_1F_1$ population, two genetic maps were constructed with 298 RFLP and 370 SSR marker loci spread over 1844 and 2210 cM, respectively (Burow et al. 2001; Foncéka et al. 2009). Considering the importance of immortal population for high-resolution mapping, several maps were prepared in the RIL populations. With the SSR, CAPS, SNP and transposon markers, 29 to 1685 loci were mapped on 8–26 LGs, map density and inter-marker distance of these map were 402–2208 cM and 0.8–18.5 cM, respectively (Jiang et al. 2007; Hong et al. 2008, 2009; Varshney et al. 2009b; Peng et al., 2010; Ravi et al. 2011; Sarvamangala et al. 2011; Hong et al. 2010; Khedikar et al. 2010; Sujay et al. 2012; Gautami et al. 2012b; Qin et al. 2012; Mondal et al. 2012; Zhou et al. 2014; Zhao et al. 2016). Recently, using NGS-based ddRADseq technique, Zhou et al. (2014) provided a well-saturated map for the tetraploid peanut and mapped 1685 marker loci, including 1621 SNPs and 64 SSR markers spanning a distance of 1447 cM with the average distance of 0.9 cM. Use of DArT and DArTseq based genotyping resulted in development of three genetic maps using $F_2$ populations with 854 loci (ICGV 07368 × ICGV 06420; Shasidhar et al. 2017), 1152 loci (ICGV 00350 × ICGV 97045; Vishwakarma et al. 2016) and 1435 loci (ICGV 06420 × SunOleic 95R; Shasidhar et al. 2017).

The first SSR-based genetic map was developed using a RIL population (TAG 24 × ICGV 86031) with 135 loci covering 1270.5 cM map distance (Varshney et al. 2009b). This genetic map was then further saturated to 191 SSR mapped loci covering 1785.4 cM genome distance (Ravi et al. 2011). Subsequently, other two SSR-based genetic maps were prepared with 56 (462.24 cM; TAG 24 × GPBD 4; Khedikar et al. 2010), and 45 marker loci (657.9 cM; TG 26 × GPBD 4; Sarvamangala et al. 2011). Later, these maps were saturated to 188 (1922.4 cM) and 181 (1963 cM) marker loci, respectively (Sujay et al. 2012). Two more RIL populations derived from the cross ICGS 44 × ICGS 76 and ICGS 76 × CSMG 84–1 were used for genetic map construction with 82 (831.4 cM) and 119 (2208.2 cM) marker loci, respectively (Gautami et al. 2012a). In addition to individual genetic maps, different maps were combined to construct integrated or consensus genetic maps. Genetic mapping information from two RIL mapping populations (TAG 24 × GPBD 4 and TG 26 × GPBD 4) segregating for foliar disease resistance

were used for constructing the first consensus map with 225 SSR loci covering total map distance of 1152.9 cM (Sujay et al. 2012). The second consensus map was developed using three RIL populations (TAG 24 × ICGV 86031, ICGS 44 × ICGS 76 and ICGS 76 × CSMG 84-1) segregating for drought tolerance related traits, and mapped 293 marker loci LGs (2840.8 cM) (Gautami et al. 2012a). In an international effort, reference consensus genetic map was prepared by using 10 RILs and one backcross (BC) populations with 897 marker loci on 20 LGs spanning a map distance of 3863.6 cM with an average map density of 4.4 cM (Gautami et al. 2012b). This consensus map was further improved by adding five more populations and achieved a dense consensus genetic map with 3693 marker loci covering 2651 cM distance (Shirasawa et al. 2013).

## 7.5 Trait Mapping Through Linkage Mapping

Initial trait mapping work started with the RAPD markers to identifying linked markers with the root-knot nematode resistance (Garcia et al. 1996; Burow et al. 1996). The RAPD markers were not preferred due to several technical problems. Nevertheless, few of these were later converted to sequence characterized amplified region (SCAR) markers for deploying in genetic and breeding studies (Chu et al. 2007). Other studies included AFLP, SSR and SNP markers for establishing an association with resistance to groundnut rosette disease and *Sclerotinia* blight, and oil quality traits namely oleic acid and linoleic acid (Herselman et al. 2004; Chenault and Maas 2006 and Lopez et al. 2000). More efforts were initiated to dissect important agronomic traits with latest NGS-based technology like GBS and array-based genotyping in peanut. Nonetheless, to achieve the sustainable yield of crop breeding for several biotic and abiotic stresses is mandatory. Handful genomics tools and techniques provided breeders a new way to dissect useful QTLs/genes leading to their deployment in breeding (Janila et al. 2016b; Pandey et al. 2016a). In total, 46 major QTLs were identified for several biotic stresses with the phenotypic variation explained (PVE) 10.05–82.96%, 59 for the abiotic stresses and their related traits with the PVE range of 10.0–22.24%, 50 major QTLs for agronomic and yield component traits with the PVE range of 10.1–33.36% and 50 major QTLs for other morphological traits with 10.0–28.2% PVE and 60 major QTLs for seed and oil quality traits with the PVE range of 10.2–45.63% (Table 7.2).

In case of peanut, several biotic stresses affect yield and quality adversely including rust, ELS and LLS, nematode, GRD, TSWV, bacterial wilt, *Sclerotinia minor*, *Aspergillus* and aflatoxin contamination. For rust resistance, 18 major QTLs were reported with 10.68–82.96% PVE (Khedikar et al. 2010; Sujay et al. 2012; Mondal et al. 2014a; Leal-Bertioli et al. 2015). Similarly, 15 major QTLs for LLS resistance with the PVE range of 10.27–67.98%, 4 major QTLs for GRV resistance with the PVE range of 10.05–76.1%, 5 major QTLs for TSWV with the PVE range of 10.64–35.8%, 2 major QTLs for bacterial wilt resistance with the PVE range of 12–22% and 13 major QTLs for nematode resistance with the PVE range of 11.9–22.1% (Herselman et al. 2004; Liang et al. 2009; Sujay et al. 2012; Qin et al. 2012; Wang et al. 2013; Burow et al. 2014; Zhao et al. 2016; Leal-Bertioli et al. 2016). For aflatoxin contamination (AC), so far only three major QTLs were identified with 10.5–22.7% PVE (Liang et al. 2009). In addition to this, Mondal et al. (2014b) identified QTLs for the Bruchid resistance component traits.

Among the abiotic stresses, heat and terminal drought are the two major stress factors causing severe yield loss and quality deterioration of the produce in peanut in addition to other factors such as sodic and acidic nature of soil, micronutrients deficiency (Zinc, Iron) and aluminium toxicity (Janila and Nigam 2013). In peanut, ICRISAT with the research partners has done pioneer work to identify linked markers for drought tolerance related traits. In this context, major QTLs were identified successfully for transpiration (5 QTLs), transpiration efficiency (4 QTLs), carbon discrimination ratio (1 QTL), specific leaf area (6 QTLs), leaf area (1 QTL), SPAD chlorophyll

**Table 7.2** List of major QTLs identified for important traits in peanut

| S. N. | Traits studied | Major QTLs identified | Phenotypic variation explained (%) | References |
|---|---|---|---|---|
| *Biotic stress resistance* | | | | |
| 1 | Late leaf spot (LLS) | 20 | 10.3–68.0 | Sujay et al. (2012), Wang et al. (2013), Pandey et al. (2016b, 2017b) |
| 2 | Leaf rust | 18 | 10.7–83.0 | Khedikar et al. (2010), Sujay et al. (2012), Mondal et al. (2014a), Leal-Bertioli et al. (2015), Pandey et al. (2016b) |
| 3 | Resistance to *Aspergillus flavus* invasion | 3 | 10.5–22.7 | Liang et al. (2009) |
| 4 | Aphid vector of groundnut rosette disease | 4 | 10.1–76.1 | Herselman et al. (2004) |
| 5 | Resistance to tomato spotted wilt virus (TSWV) | 6 | 10.6–35.8 | Qin et al. (2012), Wang et al. (2013), Pandey et al. (2017b) |
| 6 | Root-knot nematode | 13 | 11.9–22.1 | Burow et al. (2014), Leal-Bertioli et al. (2016) |
| 7 | Bacterial wilt (BW) | 4 | 12.0–22.0 | Peng et al. (2010), Zhao et al. (2016) |
| 8 | Bruchid resistance component traits | 10 | | Mondal et al. (2014b) |
| *Abiotic stress tolerance* | | | | |
| 9 | Transpiration (T) | 5 | 10.3–18.2 | Varshney et al. (2009b), Ravi et al. (2011), Gautami et al. (2012a) |
| 10 | Transpiration efficiency (TE) | 4 | 12.3 | Ravi et al. (2011), Gautami et al. (2012a) |
| 11 | Specific leaf area (SLA) | 9 | 11.0–20.3 | Varshney et al. (2009b), Ravi et al. (2011) |
| 12 | Leaf area (LA) | 1 | 11.5 | Ravi et al. (2011) |
| 13 | SPAD chlorophyll metre reading (SCMR) | 17 | 10.6–31.2 | Varshney et al. (2009b), Ravi et al. (2011) |
| 14 | Biomass | 3 | 15.6–20.3 | Ravi et al. (2011) |
| 15 | Canopy conductance (ISC) | 3 | 11.9–22.2 | Ravi et al. (2011) |
| 16 | Total dry matter (TDM) | 1 | 22.4 | Gautami et al. (2012a) |
| 17 | Harvest index | 1 | 18.1 | Fonceka et al. (2012) |
| 18 | Hundred pod weight | 2 | 15.0–17.0 | Fonceka et al. (2012) |
| 19 | Hundred seed weight | 2 | 12.4–14.9 | Fonceka et al. (2012) |
| 20 | Haulm weight | 2 | 13.5–17.5 | Fonceka et al. (2012) |
| 21 | Pod number | 2 | 9.6–12.6 | Fonceka et al. (2012) |
| 22 | Total biomass | 2 | 11.0–16.6 | Fonceka et al. (2012) |

(continued)

**Table 7.2** (continued)

| S. N. | Traits studied | Major QTLs identified | Phenotypic variation explained (%) | References |
|---|---|---|---|---|
| 23 | Stress tolerance indices (STI)—Hundred pod weight | 2 | 13.9–16.8 | Fonceka et al. (2012) |
| 24 | STI—Hundred seed weight | 2 | 15.5–16.2 | Fonceka et al. (2012) |
| 25 | STI—Haulm weight | 2 | 16.4–17.1 | Fonceka et al. (2012) |
| 26 | STI—Pod number | 2 | 10.4–19.4 | Fonceka et al. (2012) |
| 27 | STI—Pod weight | 1 | 12.3 | Fonceka et al. (2012) |
| 28 | STI—Seed number | 6 | 11.0–26.0 | Fonceka et al. (2012) |
| 29 | STI—Seed weight | 5 | 11.5–15.2 | Fonceka et al. (2012) |
| 30 | STI—Total biomass | 2 | 10.8–20.1 | Fonceka et al. (2012) |
| *Agronomic and yield component traits* | | | | |
| 31 | Shoot dry weight (ShDW) | 2 | 14.4–22.1 | Gautami et al. (2012a) |
| 32 | Haulm weight | 2 | 10.4–36.1 | Ravi et al. 2011, Fonceka et al. (2012) |
| 33 | Harvest index | 2 | 11.0–40.1 | Gautami et al. (2012a), Fonceka et al. (2012) |
| 34 | Pod mass/plant | 3 | 13.1–18.3 | Liang et al. (2009) |
| 35 | Mature pods/plant | 2 | 11.9–12.3 | Liang et al. (2009) |
| 36 | Pod number | 1 | 14.2 | Fonceka et al. (2012) |
| 37 | Number of branches | 3 | 10.2–17.3 | Liang et al. (2009) |
| 38 | Number of fruit branches | 1 | 17.5 | Liang et al. (2009) |
| 39 | Height of main axis | 3 | 10.3–12.8 | Liang et al. (2009) |
| 40 | Stem diameter | 2 | 10.4–24.1 | Liang et al. (2009) |
| 41 | Leaf length, width and length/width ratio | 7 | 12.4–18.9 | Liang et al. (2009) |
| 42 | Length of main stem | 2 | 15.7–19.2 | Shirasawa et al. (2012b) |
| 43 | Length of the longest branch | 2 | 14.2–21.1 | Shirasawa et al. (2012b) |
| 44 | Number of branches | 1 | 15.6 | Shirasawa et al. (2012b) |
| 45 | Weight of plant | 1 | 11.8 | Shirasawa et al. (2012b) |
| 46 | Weight of mature pod per a plant | 1 | 28.1 | Shirasawa et al. (2012b) |
| 47 | Weight of seeds | 1 | 19.1 | Shirasawa et al. (2012b) |
| 48 | Yield parameters | 5 | 10.1–17.7 | Selvaraj et al. (2009) |
| 49 | Hundred pod weight | 2 | 15.1–20.6 | Fonceka et al. (2012) |
| 50 | Hundred seed weight | 2 | 15.7–16.3 | Fonceka et al. (2012) |
| 51 | Pod weight | 1 | 11.7 | Fonceka et al. (2012) |
| 52 | Shell weight | 1 | 12.6 | Fonceka et al. (2012) |

(continued)

**Table 7.2** (continued)

| S. N. | Traits studied | Major QTLs identified | Phenotypic variation explained (%) | References |
|---|---|---|---|---|
| 53 | Seed number | 1 | 14.5 | Fonceka et al., 2012 |
| 54 | Seed weight | 1 | 11.0 | Fonceka et al. (2012) |
| 55 | Total biomass | 1 | 13.2 | Fonceka et al. (2012) |
| *Other morphological traits* | | | | |
| 56 | Flowering date | 1 | 19.5 | Shirasawa et al. (2012a, b) |
| 57 | Angle of branch | 2 | 11.9–23.2 | Shirasawa et al. (2012a, b) |
| 58 | Constriction of pod | 7 | 10.0–23.9 | Shirasawa et al. (2012a, b), Fonceka et al. (2012) |
| 59 | Pod beak | 5 | 11.6–17.4 | Fonceka et al. (2012) |
| 60 | Pod length | 5 | 20.5–28.2 | Shirasawa et al. (2012b), Chen et al. (2016b) |
| 61 | Thickness of pod | 1 | 21.7 | Shirasawa et al. (2012a, b) |
| 62 | Pod width | 8 | 15.2–25.5 | Shirasawa et al. (2012a, b), Fonceka et al. (2012), Chen et al. (2016b) |
| 63 | Seed length | 8 | 11.2–20.8 | Fonceka et al. (2012), Chen et al. (2016b) |
| 64 | Seed width | 4 | 14.2–23.7 | Fonceka et al. (2012), Chen et al. (2016b) |
| 65 | Growth habit | 5 | 13.9–17.3 | Fonceka et al. (2012) |
| 66 | Main stem height | 4 | 10.0–26.7 | Fonceka et al. (2012) |
| *Seed and oil quality* | | | | |
| 67 | Oil content | 6 | 10.2–14.2 | Selvaraj et al. (2009), Sarvamangala et al. (2011), Pandey et al. (2014c) |
| 68 | Protein content | 3 | 10.2–13.4 | Liang et al. (2009), Sarvamangala et al. (2011) |
| 69 | Carbon discrimination ratio | 1 | 12.2 | Ravi et al. (2011) |
| 70 | Oleic acid | 9 | 10.7–38.4 | Pandey et al. (2014c) |
| 71 | Linoleic acid | 8 | 12.0–39.5 | Pandey et al. (2014c) |
| 72 | Oleic/linoleic (O/L) acid ratio | 3 | 10.8–45.6 | Pandey et al. (2014c) |
| 73 | Palmitic acid | 6 | 10.6–37.4 | Wang et al. (2014) |
| 74 | Stearic acid | 6 | 17.8–40.57 | Wang et al. (2014) |
| 75 | Arachidic acid | 4 | 28.3–36.9 | Wang et al. (2014) |
| 76 | Gadoleic acid | 9 | 11.2–26.1 | Wang et al. (2014) |
| 77 | Behenic acid | 2 | 12.4–13.6 | Wang et al. (2014) |
| 78 | Lignoceric acid | 3 | 10.0–12.6 | Wang et al. (2014) |
| 79 | Total phenolics | 1 | 12.5 | Mondal et al. (2015) |
| 80 | Total flavonoids | 5 | 25.0–67.0 | Mondal et al. (2015) |
| 81 | DPPH radical scavenging | 4 | 11.5–33.0 | Mondal et al. (2015) |

metre reading (SCMR) (12 QTLs), biomass (3 QTLs), shoot dry weight (2 QTLs), haulm weight (1 QTL), harvest index (1 QTL), canopy conductance (3 QTLs) and total dry matter (1 QTL) (Varshney et al. 2009b; Ravi et al. 2011; Gautami et al. 2012a). The another study while dissecting drought tolerance traits Fonceka et al. (2012) identified two major QTLs each for 100 pod weight, 100 seed weight, haulm weight, pod number, total biomass, STI-100 pod weight, STI-100 seed weight, STI-haulm weight, STI-pod number, STI-total biomass, one major QTL each for STI-pod weight, STI-seed number and STI-seed weight. Although, above-mentioned studies provided preliminary idea on the probable genomic regions showing association with drought tolerance related traits, but still no linked marker has been validated so far which can be deployed in breeding. Nevertheless, lots of genetic populations have been developed while conducting above studies. Generation of multi-location and multi-replicated phenotyping and high throughput genotyping data on these genetic populations will facilitate high-resolution trait mapping and identification of linked markers for drought tolerance related traits.

## 7.6 Oil and Nutritional Quality

In most populated countries like China and India, peanut is mainly crushed for oil extraction while it mainly serve as table purpose and preferred for low oil content in other countries like United States of America. For oil content, 6 major QTLs were identified with the PVE ranging from 10.2 to 14.18% (Selvaraj et al. 2009; Sarvamangala et al. 2011; Pandey et al. 2014b). For protein content, to date only two workers namely Liang et al. (2009) and Sarvamangala et al. (2011) separately reported three major QTLs with PVE ranging from 10.7% to 13.4%.

Peanut oil contains unsaturated fatty acids (UFA), i.e. oleic and linoleic acid and saturated fatty acids (SFA), i.e. palmitic acid, stearic acid, arachidic acid, behenic acid, lignoceric acid and gadoleic acid. The UFAs were further categorized into monounsaturated fatty acid (MUFA) such as oleic acid (C18:1), and polyunsaturated fatty acid (PUFA), i.e. linoleic acid (C18:2). Oleic acid is known to diminish the risk of cardiovascular disease (CVD) by decreasing the levels of serum low-density lipoproteins (LDL) cholesterol and preserving the levels of high-density lipoproteins (HDL). The ratio of UFAs in peanut oil comprises 80% while remaining 20% accounted by SFAs, apart from this UFA is also very high in peanut oil in comparison to the butter, coconut oil and palm oil that bestow peanut oil as a healthier food as a consuming oil (Johnson et al. 2009). With the addition of a double bond to C18:1 (oleic acid), oleic acid converts into linoleic acid, and this reaction is catalyzed by the fatty acid desaturase (FAD) enzyme (Ray et al. 1993). Genetics studies revealed that the high oleic acid is controlled by two homozygous recessive mutant alleles of FAD2A and FAD2B genes. The first study of reporting QTLs for oleic acid, linoleic acid and oleic/linoleic (O/L) ratio other than FAD2A and FAD2B reported 20 major QTLs with the PVE range 10.71–45.63 (Pandey et al. 2014b). Another study reported 30 major QTLs for saturated fatty acids (Wang et al. 2015). In addition to this, ten major QTLs were reported with the PVE% range 12.5, 25–67 and 11.5–33 for the total phenolics, total flavonoids and DPPH radical scavenging, respectively by the Mondal et al. (2015). The linked markers are available for use in breeding to improve the fatty acid profiles in peanut.

## 7.7 Agronomic and Morphological Traits

Enhancing the pod yield has been the main goal since the first day of breeding and will remain the main goal even in future seeing the unprecedented population growth globally. In this context, efforts were made to identify the QTLs associated with yield and yield component traits. So far, a total 50

QTLs were reported for yield and yield component traits with PVE range 10.1–40.1% (Selvaraj et al. 2009; Liang et al. 2009; Fonceka et al. 2012; Shirasawa et al. 2012b). Major QTLs for flowering date, angle of branch, pod characteristics such as constriction, beak, length, thickness and width, seed width and length, pod width and length, growth habit and main stem height were identified in three separate studies (Fonceka et al. 2012; Shirasawa et al. 2012b; Chen et al. 2016b). In addition to bi-parental populations, genetic populations involving multiple parents (such as MAGIC and NAM) have been developed by ICRISAT for conducting high-resolution mapping using multi-location and multi-replicated phenotyping and high throughput genotyping data on these genetic populations.

## 7.8 Trait Mapping Through Linkage Disequilibrium/Association Mapping

Bi-parental populations have limitations for being able to provide allelic variation for few traits and enable to dissect a small fraction of the probable alleles through linkage mapping. Furthermore, genetic resolution of QTL mapping often remains limited in a range of 10–30 cM due to confined number of meiotic events that are captured in a bi-parental mapping population (Zhu et al. 2008). Globally, availability of large number of peanut germplasm provided opportunity to think out of box and utilize this germplasm in trait mapping using association mapping approach. To exploit these available set of germplasm, association mapping approach for trait mapping is very promising (Gupta et al. 2014). Association mapping can be categorized in two categories candidate gene-based association (CGAS) and genome-wide association mapping (GWAS) (Zhu et al. 2008).

Phenotyping for quality traits and genotyping of the US 'Mini Core Collection' with 81 SSR markers identified two functional SNP markers for two fatty acid desaturase (*FAD2* for oleic acid, linoleic acid and oleic-to-linoleic ratio (Wang et al. 2011). Subsequently in another study, marker-trait associations (MTAs) with low phenotypic variation (1.05–4.81% PVE) for 15 agronomic traits were identified in in Chinese 'Mini-Core Collection' (Jiang et al. 2014). Recently, at ICRISAT in order to identify MTAs, phenotyping and genotyping data (4597 polymorphic DArT loci and 154 SSR loci) generated on 'Reference Set' developed by ICRISAT comprising of 300 accessions were used for association analysis and identified 524 highly significant markers with 5.8–90.1% PVE for 36 traits (Pandey et al. 2014b). Recently, 'Reference Set' with 300 accessions, US 'Mini Core Collection' containing 112 accessions, and Chinese 'Mini Core Collection' with 298 accessions, an endeavour has been initiated at International level to generate high throughput genotyping data in addition to precise phenotyping to conduct high-resolution association analysis for identifying linked markers and accessions with superior alleles for use in the breeding programmes.

## 7.9 Advanced Backcross (AB) QTL Mapping

Wild species of peanut is a reservoir of useful genes and alleles for biotic and abiotic stresses (Upadhyaya et al. 2012). These genes were untouched and unexploited throughout the course of evolution and domestication. Despite owns boastful wild germplasm, just few (1.1%) were used to develop advanced breeding lines (Sharma et al. 2013). Recently, through remarkable attempts, some elite cultivars with multiple disease resistance were released from ICRISAT and USA (Sharma et al. 2013; Burow et al. 2013). Similar to other polyploid species, continuous domestication of cultivated peanut narrowed the genetic diversity which impose a genetic bottleneck. Since direct introgression of the useful genes from the wild species to cultivated gene pool is very difficult, the synthetic amphidiploid were used as bridge species to defeat the reproductive barriers between the wild diploids and the cultivated species. At ICRISAT, 17 new synthetic amphiploids and autotetraploids populations were developed to broadening genetic

base (Mallikarjuna et al. 2012; Shilpa et al. 2013). These new synthetics were also reported as resistance to late leaf spot and peanut bud necrosis disease.

To save the time and enhance the accuracy during identification and introgression of useful alleles from wild to cultivated genotypes, molecular markers proved as a very handful tool. In this context, strategy like advanced-backcross QTL (AB-QTL) has been suggested by the Tanksley et al. (1996). In this strategy, identification of linked markers as well as introgression of trait of interest can be done simultaneously. In this direction, 110 QTLs were identified for several traits including physiological, agronomic, morphological traits and resistance to the root-knot nematode (Fonceka et al. 2012; Burow et al. 2014). Further, at ICRISAT, two AB-QTL populations were developed and phenotyped for several disease resistance traits. The genotyping data (DArT markers) and phenotyping data were analysed leading to identification of QTLs for rust and LLS resistance (Varshney et al. 2013). Two of the above-mentioned synthetics namely ISATGR 278-18 (*A. duranesis* × *A. batizocoi*) and ISATGR 5B (*A. magna* × A. *batizocoi*) were used to introgress foliar disease resistance in five elite Indian peanut varieties namely ICGV 91114, ICGS 76, ICGV 91278, JL 24, and DH 86 using backcross breeding approach (Kumari et al. 2014). In addition to disease resistance, these lines have also shown range of variation for other important morphological and agronomic traits.

## 7.10 Next-Generation Genetic Populations for High-Resolution Trait Mapping

With the advancements in the NGS technology which can generate huge genomic sequence data in very short time. Several analysis softwares/tools have become available in public domain for analysing large data sets. These developments have allowed us to develop even more complex and larger genetic populations to perform high-resolution trait mapping. Some of these important populations include MAGIC, NAM and recombinant inbred advanced intercross line (RIAIL) populations (Morrell et al. 2012). It is important to note that MAGIC population facilitate increased recombination events by making multiple crosses using multiple parents to create highly diverse genetic population for conducting high-resolution genetic mapping (Cavanagh et al. 2008). Analysis of such population has remained challenge due to presence of multiple alleles coming from different founder parental genotypes in the population. Nevertheless, a whole-genome average interval mapping (WGAIM) approach has been developed recently for conducting QTL analysis (Verbyla et al. 2014). This approach is very promising for conducting high-resolution trait mapping for several traits simultaneously. Recently, at ICRISAT two MAGIC populations for (aflatoxin resistance and drought tolerance) and two MAGIC population for agronomic traits have been developed. NAM is another promising approach to dissect the genetic basis of complex traits by capturing genetic diversity of selected diverse parents (founders). Most importantly, the NAM population has higher power QTL detection as compared with bi-parental mapping populations (Yu et al. 2008; McMullen et al. 2009). In peanut, two NAM populations, i.e. one each for Spanish (ICGV 91114 and 22 testers) and Virginia type (ICGS 76 and 21 testers) were developed (Varshney 2016; Pandey et al. 2016a). The development of MAGIC and NAM populations have given birth to a new method of trait mapping called joint linkage-association mapping (JLAM). These populations can be used to conduct linkage as well as association mapping and are very useful for conducting high-resolution mapping (Gupta et al. 2014).

## 7.11 Emerging NGS-Based Trait Mapping Strategies

It has been discovered that a high level of resolution can be achieved with the help of high-density genotyping by using NGS methods

(Huang et al. 2009) and mapping by sequencing (Schneeberger and Weigel 2011). Recently, few trait mapping approaches demonstrated speedy detection of genomic regions and candidate genes controlling the targeted traits such as MutMap, QTL-seq and BSR-seq.

Abe et al. (2012) identified successfully causal mutations for pale green leaves and semi-dwarfism in rice using MutMap approach. In this approach, whole-genome re-sequencing (WGRS) is performed for the pooled DNA samples from a $F_2$ segregating progeny of a cross between a mutant and its wild type (WT). The concept of SNP identification is based on the SNP index and the sequence data of bulked DNA is aligned with the reference sequence. The SNPs with sequence reads containing only of the mutant sequences (SNP index = 1) are considered to be linked to the causal SNP for the mutant phenotype. The MutMap is theoretically similar to some of other related methods such as SHOREmap (Schneeberger et al. 2009) and next-generation mapping (Austin et al. 2011). The same group has updated MutMap to MutMap plus where the same concept of identification of causative SNP for the mutant phenotype can be achieved without crossing the mutant with WT line. Therefore, the DNA of $M_3$ progenies with extreme phenotypes are bulked to get the SNP index (Fekih et al. 2013). To overcome the difficulty of mutations in the missing genomic regions from the reference (gap) genome when the reference genome is aligned to the re-sequenced lines, Takagi et al. (2013a) has proposed MutMap-Gap strategy where MutMap is used to identify the causal SNP followed by de novo assembly, alignment, and identification of the causal mutation within the genome gaps. In peanut, this strategy can be implemented too for the identification of agronomic traits of interest.

In the QTL-seq technique, MutMap strategy was conceptually integrated to the normal $F_2$ and RIL population (Takagi et al. 2013b). The principle involves a combination of bulked segregant analysis and whole genome re-sequencing for rapid identification of agronomically important QTLs. After alignment of the sequence with reference sequence, SNP index is derived to narrow down to the causal SNP (Takagi et al. 2013b). This strategy has been used to identify an early flowering QTL in cucumber (Lu et al. 2014). In peanut, by using QTL-seq approach three for rust resistance and one for LLS resistance allele-specific diagnostic markers were identified (Pandey et al. 2016b). These results prove the usefulness of QTL-seq approach for the rapid and precise and identification of candidate genomic regions and development of diagnostic markers for breeding applications.

BSR-sequencing is one of the potential methods where the candidate genes/markers associated with the trait can be identified at the expression level. DNA-based BSA requires access to quantitative genetic markers that are polymorphic in the mapping population. The modification over BSA method, BSR-Seq makes use of RNA-Seq reads to efficiently map genes even in populations for which no polymorphic markers have been previously identified (Liu et al. 2012). In this approach, it is possible to conduct de novo SNP discovery and quantitatively genotype the bulks from extreme phenotype or mutants using appropriate computational tools. This is relatively new technique that is less explored in plants and there are no published reports yet in peanut. As this approach uses the expression data into consideration, there is advantage of identifying probable candidate genes to dissect important traits.

## 7.12 Molecular Breeding for the Disease Resistance and Oil Content and Quality Traits

The identification of molecular markers linked to desirable traits in peanut has provided the pace to the peanut improvement programs using molecular breeding. Two molecular breeding approaches namely marker-assisted backcrossing (MABC) and marker-assisted selection (MAS) facilitate transfer QTLs/gene from source genotype to elite recipient cultivars. The MABC

and MAS approaches are very precise in selection at very initial stage of the plant through the trait linked markers. Additionally, MABC approach shortens the generation to achieve higher recurrent parent genome recovery as compared to conventional breeding methods. Some of the successful examples of MABC/MAS application and their output have been discussed below.

At the earliest through MABC approach, Simpson et al. (2003) developed Nematode resistance lines and registered as 'NemaTAM' variety. Chu et al. (2011) pyramided high oleic acid and nematode resistance in cultivated peanut and also developed the CAPS markers for *ahFAD2A* and *ahFAD2B* mutant alleles responsible for oil quality traits. The South African peanut cultivars were improved for the high oleic acid trait through MAS (Mienie and Pretorius, 2013). In Indian continent, rust and LLS are the major foliar fungal diseases of peanut causing 40–70% losses in pod yield. Most of the popular cultivars in major growing state viz. Maharashtra, Karnataka, Tamil Nadu and Andhra Pradesh have been reported susceptible to rust and LLS. With an objective to breed resistant varieties for foliar disease resistance, Varshney et al. (2014) introgressed one major QTL each for rust resistance and LLS resistance conferring >80% and 67.98% PVE, respectively, in the popular varieties namely ICGV 91114, JL 24 and TAG 24. Furthermore recently, Janila et al. (2016c) evaluated these selected introgression lines at three locations including disease hot spots regions of India. The reason was to assure the expression of resistance governed by the QTL region, as different factors viz. genotype background, environment and genotype × environment interactions work behind this. Resultantly, six best ILs namely ICGV 13192, ICGV 13193, ICGV 13200, ICGV 13206, ICGV 13228 and ICGV 13229 were picked with 39–79% higher mean pod yield and 25–89% higher mean haulm yield in comparison to their respective recurrent parents. Pod yield increase was contributed by increase in seed mass and number of pods per plant. The most interesting result was combining short maturity duration together with foliar disease resistance through MABC approach which was not earlier achieved through conventional breeding approaches. Similarly, for improving quality traits, MABC/MAS approaches were used to improve three major fatty acids namely oleic, linoleic and palmitic acids by transferring two mutant alleles from donor 'SunOleic 95R' in three Indian elite varieties namely ICGV 06110, ICGV 06142 and ICGV 06420 (Janila et al. 2016a). Now the efforts are underway to combine foliar disease resistance and oil quality through marker-assisted gene pyramiding approach.

## 7.13 Conclusion

Peanut is a crop of global importance and is an essential component of human food basket. This crop has been lacking optimal genomic resources to improve the breeding efficiency for achieving higher genetic gain in less time. The year '2016' has been very good as genome assemblies for both the diploid ancestors of cultivated peanut were made available. Last couple of years were also good in the context of developing several genetic populations and preliminary genetic mapping and trait mapping. It is equally important that now peanut is also witnessing development of high throughput genotyping platforms and high-resolution multi-parent mapping populations. The availability of such resources will further accelerate development and deployment of genomic resources targeting peanut genetic improvement.

**Acknowledgements** We would like to express our appreciation to the financial support from the Peanut Foundation, MARS Inc., the Georgia Peanut Commission and the U.S. National Peanut Board, Bill & Melinda Gates Foundation (Tropical Legumes I, II & III), Department of Biotechnology (DBT) of Government of India and World Bank Assisted Watershed Development Project II (KWDP-II) by Government of Karnataka, India. The work reported in this article was undertaken as a part of the CGIAR Research Program on Grain Legumes. ICRISAT is a member of the CGIAR.

## References


Abe A, Kosugi S, Yoshida K et al (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. Nat Biotechnol 30:174–179

Alpert KB, Tanksley SD (1996) High-resolution mapping and isolation of a yeast artificial chromosome contig containing fw2.2: a major fruit weight quantitative trait locus in tomato. Proc Natl Acad Sci USA 93:15503–15507

Austin RS, Vidaurre D, Stamatiou G et al (2011) Next generation mapping of Arabidopsis genes. Plant J 67:715–725

Beissinger TM, Hirsch CN, Sekhon RS, Foerster JM, Johnson JM, Muttoni G et al (2013) Marker density and read depth for genotyping populations using genotyping-by-sequencing. Genetics 193:1073–1081. doi:10.1534/genetics.112.147710

Bertioli SCM, José ACVF, Alves-Freitas DMT, Moretzsohn MC, Guimarães PM, Nielen S, Vidigal BS et al (2009) Identification of candidate genome regions controlling disease resistance in *Arachis*. BMC Plant Biol 9(1):112

Bertioli DJ, Ozias-Akins P, Chu Y, Dantas KM, Santos SP, Gouvea E, Guimarães PM, Leal-Bertioli SC, Knapp SJ, Moretzsohn MC (2014) The use of SNP markers for linkage mapping in diploid and tetraploid peanuts. G3 (Bethesda) 4(1):89–96

Bertioli DJ, Cannon SB, Froenicke Lutz, Huang G et al (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nature. doi:10.1038/ng.3517

Burow MD, Simpson CE, Paterson AH, Starr JL et al (1996) Identification of peanut (*Arachis hypogaea* L.) RAPD markers diagnostic of root-knot nematode (Meloidogyne arenaria (Neal) Chitwood) resistance. Mol Breed: new strat plant improv 2:369–379

Burow MD, Simpson CE, Starr JL, Paterson AH (2001) Transmission genetics of chromatin from a synthetic amphiploid in cultivated peanut (*A. hypogaea* L.): Broadening the gene pool of a monophyletic polyploid species. Genetics 159:823–837

Burow MD, Simpson CE, Faries MW, Starr JL, Paterson A (2009) Molecular biogeographic study of recently described B- and A-genome *Arachis* species, also providing new insights into the origins of cultivated peanut. Genome 52:107–119

Burow MD, Leal-Bertioli SC, Simpson CE, Ozias-Akins P, Chu Y, Denwar NN, Chagoya J, Starr JL, Moretzsohn MC, Pandey MK et al (2013) Marker-assisted selection for biotic stress resistance in peanut. Transl Genomics Crop Breed: 125–150

Burow MD, Starr JL, Park C, Simpson CE, Paterson AH (2014) Introgression of homeologous quantitative trait loci (QTLs) for resistance to the root-knot nematode [*Meloidogyne arenaria* (Neal) Chitwood] in an advanced backcross-QTL population of peanut (*Arachis hypogaea* L.). Mol Breed 34(2):393–406

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 X, Li H, Pandey MK et al (2016a) Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis and allergens. PNAS 113(24):6785–6790. doi:10.1073/pnas.1600899113

Chen W, Jiao Y, Cheng L et al (2016b) Quantitative trait locus analysis for pod- and kernel-related traits in the cultivated peanut (*Arachis hypogaea* L.). BMC Genet 17:25

Chenault KD, Maas AL (2006) Identification of a simple sequence repeat (SSR) marker in cultivated peanut (*Arachis hypogaea* L.) potentially associated with Sclerotinia blight resistance. Proc Am Peanut Res Educ Soc 37:24–25

Chopra R, Burow G, Farmer A, Mudge J, Simpson CE, Wilkins TA, Baring MR, Puppala N, Chamberlin KD, Burow MD (2015) Next-generation transcriptome sequencing, SNP discovery and validation in four market classes of peanut, *Arachis hypogaea* L. Mol Genet Genom 290:1169–1180

Chu Y, Holbrook CC, Timper P, Ozias-Akins P (2007) Development of a PCR-based molecular marker to select for nematode resistance in peanut. Crop Sci 47:841–845

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

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510

Dwivedi D, Johri BN (2003) Antifungals from fluorescent pseudomonads: biosynthesis and regulation. Curr Sci 85(12):1693–1703

Dwivedi SL, Bertioli DJ, Crouch JH, Valls JF, Upadhyaya HD, Fávero A, Moretzsohn M, Paterson AH (2007) Peanut. In: Kole C (ed) Genome mapping and molecular breeding in plants. Springer, Berlin, pp 115–151

Dwivedi SL, Upadhyaya HD, Stalker HT, Blair MW, Bertioli D, Nielen S, Ortiz R (2008) Enhancing crop gene pools of cereals and legumes with beneficial traits using wild relatives. Plant Breed Rev, vol 30. Wiley, New York, pp 179–280

FAOSTAT. http://faostat.fao.org/ (accessed on 10 March 2016)

Fekih R, Takagi H, Tamiru M et al (2013) MutMap +: genetic mapping and mutant identification without crossing in rice. PLoS ONE 8(7):e68529

Fonceka D, Tossim HA, Rivallan R, Vignes H et al (2012) Fostered and left behind alleles in peanut: interspecific QTL mapping reveals footprints of domestication and useful natural variation for breeding. BMC Plant Biol 12:26

Foncéka D, Tossim HA, Rivallan R et al (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

Garcia GM, Stalker HT, Schroeder E, Kochert G (1996) Identification of RAPD, SCAR, and RFLP markers tightly linked to nematode resistance genes introgressed from *Arachis cardenasii* into *Arachis hypogaea*. Genome 39(5):836–845

Garcia GM, Stalker HT, Schroeder E, Lyerly JH, Kochert G (2005) A RAPD-based linkage map of peanut based on a backcross population between the two diploid species *Arachis stenosperma* and *A. cardenasii*. Peanut Sci 32:1–8

Gautami B, Pandey MK, Vadez V et al (2012a) QTL analysis and consensus genetic map for drought tolerance traits based on three RIL populations of cultivated groundnut (*Arachis hypogaea* L.). Mol Breed 32:757–772

Gautami B, Foncéka D, Pandey MK et al (2012b) An international reference consensus genetic map with 897 marker loci based on 11 mapping populations for tetraploid groundnut (*Arachis hypogaea* L.). PLoS ONE 7(7):e41213

Gowda MVC, Motagi BN, Naidu GK, Diddimani SB, Sheshagiri R (2002) GPBD 4: A Spanish bunch groundnut genotype resistant to rust and late leaf spot. Int Arachis Newsl 22:29–32

Guo Y, Khanal S, Tang S et al (2012) Comparative mapping in intraspesific populations uncovers high degree of macrosynteny between A- and B-genome diploid species of peanut. BMC Genom 13:608

Guo B, Pandey MK, He G et al (2013) Recent advances in molecular genetic linkage maps of cultivated peanut (*Arachis hypogaea* L.). Peanut. Science 40(2):95–106

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

Gupta PK, Roy JK, Prasad M (2001) Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Curr Sci 80(4):25

Gupta PK, Kulwal PL, Jaiswal V (2014) Association mapping in crop plants: opportunities and challenges. Adv Genet 85:109

Halward T, Stalker HT, Kochert G (1993) Development of an RFLP linkage map in diploid peanut species. Theor Appl Genet 87(3):379–384

He G, Prakash CS (1997) Identification of polymorphic DNA markers in cultivated peanut (*Arachis hypogaea* L.). Euphytica 97:143–149

Herselman L (2003) Genetic variation among Southern African cultivated peanut (*Arachis hypogaea* L.) genotypes as revealed by AFLP analysis. Euphytica 133:319–327

Herselman L, Thwaites R, Kimmins FM et al (2004) Identification and mapping of AFLP markers linked to peanut (*Arachis hypogaea* L.) resistance to the aphid vector of groundnut rosette disease. Theor Appl Genet 109:1426–1433

Hong Y, Liang X, Chen X et al (2008) Construction of genetic linkage map based on SSR markers in peanut (*Arachis hypogaea* L.). Agric Sci China 7(8):915–921

Hong Y, Liang X, Chen X et al (2009) Construction of genetic linkage map in peanut (*Arachis hypogaea* L.) cultivars. Acta Agron Sin 35(3):395–402

Hong Y, Chen X, Liang X et al (2010) A SSR based composite genetic linkage map for the cultivated peanut (*Arachis hypogaea* L.) genome. BMC Plant Biol 10:17

Hong Y, Pandey MK, Liu Y, Chen X, Liu H, Varshney RK, Liang X, Huang S (2015) Identification and evaluation of single-nucleotide polymorphisms in allotetraploid peanut (*Arachis hypogaea* L.) based on amplicon sequencing combined with high resolution melting (HRM) analysis. Front Plant Sci 6. doi:10.3339/fpls.2015.01068

Hopkins MS, Casa AM, Wang T et al (1999) Discovery and characterization of polymorphic simple sequence repeats (SSRs) in peanut. Crop Sci 39:1243–1247

Huang X, Feng Q, Qian Q et al (2009) High-throughput genotyping by whole-genome resequencing. Genome Res 19:1068–1076

Husted L (1933) Cytological studies of the peanut Arachis I. Chromosome number and morphology. Cytologia 5:109117

Hyten DL, Song Q, Fickus EW et al (2010) High-throughput SNP discovery and assay development in common bean. BMC Genom 11:475

Imataka G, Arisaka O (2012) Chromosome analysis using spectral karyotyping (SKY). Cell Biochem Biophys 62(1):13–17

Iyer-Pascuzzi Anjali S, Sweeney MT et al (2007) Use of naturally occurring alleles for crop improvement. Rice Functional Genomics. Springer, New York, pp 107–147

Janila P, Nigam SN (2013) Phenotyping for groundnut (*Arachis hypogaea* L.) improvement. In: Panguluri SK, Kumar AA (eds) Phenotyping for plant breeding: applications of phenotyping methods for crop improvement, pp 129–167

Janila P, Pandey MK, Shasidhar Y et al (2016a) Molecular breeding for introgression of fatty acid desaturase mutant alleles (ahFAD2A and ahFAD2B) enhances oil quality in high and low oil containing peanut genotypes. Plant Sci 242:203–213

Janila P, Nigam SN, Pandey MK et al (2016b) Groundnut improvement: use of genetic and genomic tools. Front Plant Sci 4:23

Janila P, Pandey MK, Manohar SS et al (2016c) Foliar fungal disease-resistant introgression lines of groundnut (*Arachis hypogaea* L.) record higher pod and haulm yield in multilocation testing. Plant Breed 135:355–366

Jiang H, Chen B, Ren X et al (2007) Identification of SSR markers linked to bacterial wilt resistance of peanut with RILs. Chin J Oil Crop Sci 29(1):26–30

Jiang H, Huang L, Ren X et al (2014) Diversity characterization and association analysis of agronomic traits in a Chinese peanut (*Arachis hypogaea* L.) minicore collection. J Integr Plant Biol 56:159–169

Johnson S, Saikia N, Mathur HB, Agarwal HC (2009) Fatty acids profile of edible oils and fats in India. Centre for Science and Environment, New Delhi, pp 3–31

Khedikar YP, Gowda MV, Sarvamangala C et al (2010) A QTL study on late leaf spot and rust revealed one major QTL for molecular breeding for rust resistance in groundnut (*Arachis hypogaea* L.). Theor Appl Genet 121:971–984

Khera P, Upadhyaya HD, Pandey MK et al (2013) Single nucleotide polymorphism–based genetic diversity in the reference set of peanut (*Arachis* spp.) by developing and applying cost-effective kompetitive allele specific polymerase chain reaction genotyping assays. The Plant Genome 6:1–11

Kochert G, Stalker HT, Gimenes M et al (1996) RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). Am J Bot 83(10):1282–1291

Kumari V, Gowda MVC, Tasiwal V, Pandey MK et al (2014) Diversification of primary gene pool through introgression of resistance allele for foliar diseases from synthetic amphidiploids to cultivated groundnut (*Arachis hypogaea* L.). The Crop J 2(2–3):110–119

Lacks GD, Stalker HT (1993) Isozyme analyses of Arachis species and interspecific hybrids. Peanut Science 20(2):76–81

Leal-Bertioli SCM, Cavalcante U, Gouvea EG, Ballen-Taborda C, Shirasawa K, Guimaraes PM, Jackson SA, Bertioli DJ, Moretzsohn MC (2015) Identification of QTLs for rust resistance in the peanut wild species *Arachis magna* and the development of KASP markers for marker-assisted selection. G3-Genes Genomes. Genetics 5:1403–1413

Leal-Bertioli SC, Moretzsohn MC, Roberts PA et al (2016) Genetic mapping of resistance to *Meloidogyne arenaria* in *Arachis stenosperma*: a new source of nematode resistance for peanut. G3: Genes, Genomes, Genetics 6:377–390

Liang X, Zhou G, Hong Y et al (2009) Overview of research progress on peanut (*Arachis hypogaea* L.) host resistance to aflatoxin contamination and genomics at the Guangdong Academy of Agricultural Sciences. Peanut Sci 36:29–34

Liu S, Yeh CT, Tang HM et al (2012) Gene mapping via bulked segregant RNA-Seq (BSR-Seq). PLoS ONE 7 (5):e36406

Lopez Y, Nadaf HL, Smith OD et al (2000) Isolation and characterization of the Delta superior 1 superior 2-fatty acid desaturase in peanut (*Arachis hypogaea* L.) and search for polymorphisms for the high oleate trait in Spanish market-type lines. Theor Appl Genet 101:1131–1138

Lu H, Lin T, Klein J et al (2014) QTLseq identifies an early flowering QTL located near Flowering Locus T in cucumber. Theor Appl Genet 127:1491–1499

Luo H, Zhu H, Li Z et al (2017) Development of SSR markers and identification of major quantitative trait loci controlling shelling percentage in cultivated peanut (*Arachis hypogaea* L.) Theor Appl Genet 130:1635–1648

Macedo SE, Moretzsohn MC, Leal-Bertioli SC et al (2012) Development and characterization of highly polymorphic long TC repeat microsatellite markers for genetic analysis of peanut. BMC Res Notes 5:86

Mallikarjuna N, Jadhav DR, Reddy K et al (2012) Screening new Arachis amphidiploids, autotetraploids for resistance to late leaf spot by detached leaf technique. Eur J Plant Pathol 132:17–21

McMullen MD, Kresovich S, Villeda HS (2009) et al. Genetic properties of the maize nested association mapping population. Science 325:737–740

Mienie CMS, Pretorius AE (2013) Application of marker-assisted selection for ahFAD2A and ahFAD2B genes governing the high-oleic acid trait in South African groundnut cultivars (*Arachis hypogaea* L.). Afr J Biotechnol 12(27):4283

Milla SR (2003) Relationships and utilization of Arachis germplasm in peanut improvement. Ph.D. Thesis, North Carolina State University

Mondal S, Badigannavar AM, D'Souza SF (2012) Molecular tagging of a rust resistance gene in cultivated groundnut (*Arachis hypogaea* L.) introgressed from *Arachis cardenasii*. Mol Breed 29:467–476

Mondal S, Hande P, Badigannavar AM (2014a) Identification of transposable element markers for a rust (*Puccinia arachidis* Speg.) resistance gene in cultivated peanut. J Phytopath 162:548–552

Mondal S, Hadapad AB, Hande PA, Badigannavar AM (2014b) Identification of quantitative trait loci for bruchid (*Caryedon serratus* Olivier) resistance components in cultivated groundnut (*Arachis hypogaea* L.). Mol Breed 33:961–973

Mondal S, Phadke RR, Badigannavar AM (2015) Genetic variability for total phenolics, flavonoids and antioxidant activity of testaless seeds of a peanut recombinant inbred line population and identification of their controlling QTLs. Euphytica 204:311–321

Moretzsohn MC, Leoi L, Proite K et al (2005) A microsatellite-based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae). Theor Appl Genet 111(6):1060–1071

Moretzsohn MC, Barbosa AV, Alves-Freitas DM et al (2009) A linkage map for the B-genome of *Arachis* (Fabaceae) and its synteny to the A-genome. BMC Plant Biol 9:40

Morrell PL, Buckler ES, Ross-Ibarra J (2012) Crop genomics: advances and applications. Nat Rev Genet 13:85–96

Nagy ED, Chu Y, Guo Y et al (2010) Recombination is suppressed in an alien introgression in peanut harboring Rma, a dominant root-knot nematode resistance gene. Mol Breed 26:357–370

Nagy ED, Guo Y, Tang S et al (2012) A high-density genetic map of *Arachis duranensis*, a diploid ancestor of cultivated peanut. BMC Genomics 13: 469

Nigam SN, Rao PRDVJ, Bhatnagar-Mathur P, Sharma KK (2012) Genetic management of virus diseases in peanut, In: Janick J (ed) Plant breeding reviews vol 36. Wiley, NewYork, pp 293–356

Pandey MK, Nigam SN, Upadhyaya HD et al (2012a) Advances in Arachis genomics for peanut improvement. Biotechnol Adv 30:639–651

Pandey MK, Gautami B, Jayakumar T et al (2012b) Highly informative genic and genomic SSR markers to facilitate molecular breeding in cultivated groundnut (*Arachis hypogaea*). Plant Breed 131:139–147

Pandey MK, Guo B, Holbrook CC et al (2014a) Molecular markers, genetic maps and QTLs for molecular breeding in peanut. In: Genetics, genomics and breeding of peanuts. genetics, genomics and breeding of crop plant, CRC Press, Boca Raton, pp 79–113

Pandey MK, Upadhyaya HD, Rathore A et al (2014b) Genome-wide association studies for 50 agronomic traits in peanut using the 'reference set' comprising 300 genotypes from 48 countries of the semi-arid tropics of the world. PLoS One 9(8):e105228

Pandey MK, Wang ML, Qiao L et al (2014c) Identification of QTLs associated with peanut oil contents in RIL populations and mapping FAD2 genes and their relative contribution towards oil quality. BMC Genetics 15:133

Pandey MK, Roorkiwal M, Singh VK et al (2016a) Emerging genomic tools for legume breeding: current status and future prospects. Front Plant Sci 7:455

Pandey MK, Khan AW, Singh VK, Vishwakarma MK, Shasidhar Y, Kumar V, Garg V, Bhat RS, Chitikineni A, Janila P, Guo B, Varshney RK (2016b) QTL-seq approach identified genomic regions and diagnostic markers for rust and late leaf spot resistance in groundnut (*Arachis hypogaea* L.). Plant Biotechnol J. doi:10.1111/pbi.12686

Pandey MK, Agarwal G, Kale SM, Clevenger J, Nayak SN, Sriswathi M, Chitikineni A, Chavarro C, Chen X, Upadhyaya HD, Vishwakarma MK, Leal-Bertioli S, Liang X, Bertioli DJ, Guo B, Jackson SA, Ozias-Akins P, Varshney RK (2017a) Development and evaluation of a high density genotyping 'Axiom_Arachis' array with 58 K SNPs for accelerating genetics and breeding in groundnut. Sci Rep 7:40577. doi:10.1038/srep40577

Pandey MK, Wang H, Khera P, Vishwakarma MK, Kale SM, Culbreath AK, Holbrook CC, Wang X, Varshney RK, Guo B (2017b) Genetic dissection of novel QTLs for resistance to leaf spots and tomato spotted wilt virus in peanut (*Arachis hypogaea* L.). Front. Plant Sci 8:25. doi:10.3389/fpls.2017.00025

Pei D, Hua L, Li-na L et al (2015) Chromosome analysis of peanut (*Arachis hypogaea* L.) based on sequential GISH-FISH. Scientia Agricultura Sinica 48 (9):1854–1863

Peng W, Jiang H, Ren X et al (2010) Construction of peanut AFLP map and analysis of bacterial wilt resistant disease QTLs. Acta Agriculturae Boreali-Sinica 25(6):81–86

Peng Z, Gallo M, Tillman BL et al (2016) Molecular marker development from transcript sequences and germplasm evaluation for cultivated peanut (*Arachis hypogaea* L.). Mol Genet Genomics 291:363–381

Qin H, Feng S, Chen C et al (2012) An integrated genetic linkage map of cultivated peanut (*Arachis hypogaea* L.) constructed from two RIL populations. Theor Appl Genet 124:653–664

Raina SN, Mukai Y (1999) Genomic in situ hybridization in Arachis (Fabaceae) identifies the diploid wild progenitors of cultivated (*A. hypogaea*) and related wild (*A. monticola*) peanut species. Plant Syst Evol 214:251–262

Ravi K, Vadez V, Isobe S et al (2011) Identification of several small-effect main QTLs and large number of epistatic QTLs for drought tolerance in groundnut (*Arachis hypogaea* L.). Theor Appl Genet 122:1119–1132

Ray TK, Holly SP, Knauft DA et al (1993) The primary defect in developing seed from the high oleate variety of peanut (*Arachis hypogaea* L.) is the absence of Δ12–desaturase activity. Plant Science 91:15–21

Robledo G, Seijo JG (2010) Species relationships among the wild B genome of Arachis species (section Arachis) based on FISH mapping of rDNA loci and heterochromatin detection: a new proposal for genome arrangement. Theor Appl Genet 121:1033–1046

Robledo G, Lavia GI, Seijo G (2009) Species relations among wild Arachis species with the A genome as revealed by FISH mapping of rDNA loci and heterochromatin detection. Theor Appl Genet 118:1295–1307

Sarvamangala C, Gowda MV, Varshney RK (2011) Identification of quantitative trait loci for protein content, oil content and oil quality for groundnut (*Arachis hypogaea* L.). Field Crops Res 122:49–59

Schneeberger K, Weigel D (2011) Fast-forward genetics enabled by new sequencing technologies. Trends Plant Sci 16:282–288

Schneeberger K, Ossowski S, Lanz C et al (2009) SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat Methods 6:550–551

Seijo JG, Lavia GI, Fernandez A et al (2004) Physical mapping of the 5S and 18S-25S RRNA gene by FISH as evidence that *Arachis duranensis* and *A. ipaensis* are the wild diploid progenitors of A. hypogaea (Leguminosae). Am J Bot 91:1294–1303

Seijo JG, Lavia GI, Ferna´ndez A et al (2007) Genomic relationships between the cultivated peanut (*Arachis hypogaea*, Leguminosae) and its close relatives revealed by double GISH. Am J Bot 94:1963–1971

Selvaraj MG, Narayana M, Schubert AM et al (2009) Identification of QTLs for pod and kernel traits in cultivated peanut by bulked segregant analysis. Electron J Biotech 12(2): 3–4

Sharma S, Upadhyaya HD, Varshney RK, Gowda CL (2013) Pre-breeding for diversification of primary genepool and genetic enhancement of grain legumes. Front Plant Sci 20(4):309

Shasidhar Y, Vishwakarma MK, Pandey MK et al (2017) Molecular mapping of oil content and fatty acids using dense genetic maps in groundnut. Front Plant Sci 8:794

Shilpa K, Sunkad G, Kurella S et al (2013) Biochemical composition and disease resistance in newly synthesized amphidiploid and autotetraploid peanuts. Food and Nutr Sci 4(2):169–176

Shirasawa K, Hirakawa H, Tabata S et al (2012a) Characterization of active miniature inverted-repeat transposable elements in the peanut genome. Theor Appl Genet 124:1429–1438

Shirasawa K, Koilkonda P, Aoki K et al (2012b) In silico polymorphism analysis for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. BMC Plant Biol 12:80

Shirasawa K, Bertioli DJ, Varshney RK et al (2013) Integrated consensus map of cultivated peanut and wild relatives reveals structures of the A and B genomes of Arachis and divergence of the legume genomes. DNA Res 20(2):173–184

Simpson CE (1991) Pathways for introgression of pest resistance into *Arachis hypogaea* L. Peanut Sci 18(1):22–26

Simpson CE, Starr JL (2003) Registration of COAN'peanut. Crop Sci 41(3):918

Simpson CE, Starr JL, Church GT et al (2003) Registration of 'NemaTAM' peanut. (Registrations Of Cultivars). Crop Sci 43(4):1561–1562

Smartt J, Stalker HT (1982) Speciation and cytogenetics in Arachis. In: Pattee HE, Young CE (eds) Peanut science and technology. Am Peanut Res Ed Soc, Yoakum, TX, pp 21–49

Smartt J, Gregory WC, Gregory MP (1978) The genomes of *Arachis hypoaea*. I. Cytogenetic studies of the putative genome donors. Euphytica 27:665–675

Stalker HT (1991) A new species in the section Arachis of peanuts with D genome. Am J Bot 78:630–637

Stalker HT, Dalmacio RD (1986) Karyotype analysis and relationships among varieties of *Arachis hypogaea* L. Cytologia 58:617–629

Stalker HT, Dhesi JS, Parry D (1991) An analysis of the B genome species *Arachis batizocoi*. Plant Syst Evol 174:159–169

Subramanian V, Gurtu S, Nageswara Rao RC, Nigam SN (2000) Identification of DNA polymorphism in cultivated groundnut using random amplified polymorphic DNA (RAPD) assay. Genome 43(4):656–660

Sujay V, Gowda MV, Pandey MK et al (2012) QTL analysis and construction of consensus genetic map for foliar diseases resistance based on two RIL populations in cultivated groundnut (*Arachis hypogaea* L.). Mol Breed 32:773–788

Takagi H, Abe A, Yoshida K et al (2013a) QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J 74:174–183

Takagi H, Uemura A, Yaegashi H et al (2013b) MutMap-Gap: whole-genome resequencing of mutant F2 progeny bulk combined with de novo assembly of gap regions identifies the rice blast resistance gene Pii. New Phytol 200:276–283

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277(5329):1063–1066

Tanksley SD, Grandillo S, Fulton TM et al (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative L. pimpinellifolium. Theor Appl Genet 92:213–224

Upadhyaya HD, Sharma S, Dwivedi SL (2011) Arachis. In: Kole C (ed) Wild crop relatives: genomics and breeding resources, Legume Crops and Forages. Springer, Berlin, pp 1–19

Upadhyaya HD, Sharma S, Dwivedi SL (2012) Arachis In: Wild crop relatives:genomic and breeding resources, legume crops and forages, Springer, Berlin, pp 1–19

Varshney RK (2016) Exciting journey of 10 years from genomes to fields and markets: Some success stories of genomics-assisted breeding in chickpea, pigeonpea and groundnut. Plant Sci 242:98–107

Varshney RK, Nayak SN, May GD et al (2009a) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27:522–530

Varshney RK, Bertioli DJ, Moretzsohn MC et al (2009b) The first SSR-based genetic linkage map for cultivated groundnut (*Arachis hypogaea* L.). Theor Appl Genet 118:729–739

Varshney RK, Mohan SM, Gaur PM et al (2013) Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. Biotechnol Adv 31:1120–1134

Varshney RK, Pandey MK, Janila P et al (2014) Marker assisted introgression of a QTL region to improve rust resistance in three elite and popular varieties of peanut (*Arachis hypogaea* L.). Theor Appl Genet 127:1771–1781

Varshney RK, Singh VK, Hickey J et al (2015a) Analytical and decision support tools for genomics-assisted breeding. Trends Plant Sci S1360–1385(15):00277. doi:10.1016/j.tplants.2015.10.018

Varshney RK, Kudapa H, Pazhamala L et al (2015b) Translational genomics in agriculture: some examples in grain legumes. Crit Rev Plant Sci 34:169–194

Verbyla AP, George AW, Cavanagh CR, Verbyla KL (2014) Whole-genome QTL analysis for MAGIC. Theor Appl Genet 127:1753–1770

Vishwakarma MK, Pandey MK, Shasidhar Y et al (2016) Identification of two major quantitative trait locus for fresh seed dormancy using the diversity arrays technology and diversity arrays technology seq based genetic map in Spanish type peanuts. Plant Breed 135(3):367–375

Wang Q, Zhang X, Tang F et al (2010) Construction of genetic linkage map of peanut (*Arachis hypogaea* L.) based on SRAP markers. Chin J Oil Crop Sci 32:374–378

Wang M, Sukumaran S, Barkley NA et al (2011) Population structure and marker-trait association analysis of the US peanut (*Arachis hypogaea* L.) mini-core collection. Theor Appl Genet 123:1307–1317

Wang H, Penmetsa RV, Yuan M et al (2012) Development and characterization of BAC-end sequence derived SSRs, and their incorporation into a new higher density genetic map for cultivated peanut (*Arachis hypogaea* L.). BMC Plant Biol 12:10

Wang H, Pandey MK, Qiao L et al (2013) Genetic mapping and QTL analysis for disease resistance using F2 and F5 generation-based genetic maps derived from Tifrunner × GT-C20 in peanut (*Arachis hypogaea* L.). Plant Genome 6:3

Wang H, Khera P, Culbreath AK et al (2014) Genetic mapping and QTL analysis of disease resistance traits in peanut population Tifrunner × GT-C20. Proceedings of Advances in *Arachis* through Genomics & Biotechnology, Savannah, USA, November 11–13

Wang M, Khera P, Pandey MK et al (2015) Genetic mapping of quantitative trait loci (QTL) controlling fatty acid composition in two recombinant inbred line populations in cultivated peanuts (*Arachis hypogaea* L.). PLoS One doi:10.1371/journal.pone.0119454

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178:539–551

Zhao Y, Prakash CS, He G (2012) Characterization and compilation of polymorphic simple sequence repeat (SSR) markers of peanut from public database. BMC Res Notes 5:362

Zhao Y, Zhang C, Chen H et al (2016) QTL mapping for bacterial wilt resistance in peanut (*Arachis hypogaea* L.). Mol Breed 36:13

Zhao C, Qiu J, Agarwal G et al (2017) Genome-wide discovery of microsatellite markers from diploid progenitor species, *Arachis duranensis* and *A. ipaensis*, and their application in cultivated peanut (*A. hypogaea*). Front Plant Sci 8:1209

Zhou X, Xia Y, Ren X et al (2014) Construction of a SNP-based genetic linkage map in cultivated peanut based on large scale marker development using next-generation double-digest restriction-site-associated DNA sequencing (ddRADseq). BMC Genom 15:351

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. Plant Genome 1:5–20

# 8 History and Impact of the International Peanut Genome Initiative: The Exciting Journey Toward Peanut Whole-Genome Sequencing

Hui Wang, Xiaohong Guo, Manish K. Pandey, Xiangyun Ji, Rajeev K. Varshney, Victor Nwosu and Baozhu Guo

**Abstract**

Peanut is one of the major oilseed crops in the world and is a staple food crop for much of the world. It also faces many challenges in production and possesses many opportunities in advancing science. The U.S. Peanut Genome Initiative (PGI) was launched in 2004, and brought to the international stage in 2006 to meet these challenges through coordination of international efforts in genome research beginning with molecular marker development and the improvement of genetic map resolution and coverage. The International Peanut Genome Initiative (IPGI) was the first committed step by the global peanut research community toward meeting these needs and challenges. Ultimately, a peanut genome sequencing project was initiated in 2010 by the Peanut Genome Consortium (PGC) and the genome sequences of two diploid peanut progenitors were published in 2016. During this time, IPGI and PGC have been guiding and leading demand-driven innovations in peanut genome research and translating the information into practical research and breeding. In this chapter, we review the background and history of IPGI and its achievement in developing improved genotypes using marker-assisted breeding. We also reviewed the development of peanut populations for high-resolution genetic and trait mapping, highlighting the transition to and preparation for next-generation, multi-parental genetic mapping populations from individual bi-parental populations.

H. Wang · X. Guo · X. Ji · B. Guo (✉)
USDA-ARS, Tifton, GA, USA
e-mail: baozhu.guo@ars.usda.gov

H. Wang · X. Guo · X. Ji
University of Georgia, Tifton, GA, USA

X. Guo
Heilongjiang Bayi Agricultural University, Daqing, China

M.K. Pandey · R.K. Varshney
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India

X. Ji
Ecological Environment Protection Research Institute, Shanghai Academy of Agricultural Sciences, Shanghai, China

V. Nwosu
Plant Science Program Research and Development, MARS Chocolate North America, Hackettstown, NJ, USA


R.K. Varshney et al. (eds.), *The Peanut Genome*,
Compendium of Plant Genomes, DOI 10.1007/978-3-319-63935-2_8

## 8.1 Introduction

A major milestone in biological science was the sequencing of the human genome which provided fundamentally novel methods of studying the human body (Lander et al. 2001; Subramanian et al. 2001; Venter et al. 2001). Likewise, plant genome sequencing is impacting our understanding of crops and their interactions with the environment. The complete decoding of the three billion letter human genetic codes marked an important milestone in biomedical research, suggesting that the human genome may contain fewer than the expected 50–100,000 genes (Lander et al. 2001). No matter how many genes are encoded in the human genome, only a fraction of them are expressed at any given time in any given cell within the human body. This is also true in the plant genome. To better understand plant genetic improvement of crop yields and plant responses to stress, more information is needed on the dynamics of gene activities in plants, and how their expression is controlled in the context of a cell as a function of time and space. By 2050, the human population on the earth is expected to reach nine billion (Nature Editorial 2010), with demands for food, feed, and fiber continuing to grow. Therefore, there is an urgent need to develop new technology to produce improved cultivars of crop plants, such as peanuts, that contribute toward feeding the increasing global population. Advances in food production will also require greater efforts in agricultural research to increase crop yield with improved genetics for plant protection from biotic and abiotic stresses.

Peanut (*Arachis hypogaea* L.) is one of the major economically important legumes cultivated worldwide for its ability to grow in semi-arid environments with relatively low inputs of chemical fertilizers. On a global basis, peanut is also a major source of protein and vegetable oil for human nutrition, containing about 28% protein, 50% oil, and 18% carbohydrates. Peanut is cultivated in more than 100 countries in Asia, Africa, and the Americas, grown mostly by resource-limited farmers of the semi-arid regions. India and China together produce almost two-thirds of the world's peanuts, and the U.S. produces about 6% (Guo et al. 2012).

Farmers face many challenges to increasing peanut productivity. Low productivity of peanut in several countries is ascribed to several stress factors including biotic and abiotic stresses in the cultivation of the crop (Khedikar et al. 2010; Pandey et al. 2014a, b, c). Among the biotic stresses, diseases are the major constraints that limit peanut productivity including yield and quality. Major peanut diseases include early leaf spot (ELS, *Cercospora arachidicola*), late leaf spot (LLS, *Phaeoisariopsis personata* Berk. and Curt.), rust (*Puccinia arachidis* Speg.), *Peanut mottle virus* (PMV), *Groundnut rosette virus* (GRV), *Tomato spotted wilt virus* (TSWV), and root-knot nematode (Pandey et al. 2014a, b, c). Rust, stem rot (*Sclerotium rolfsii*), collar rot (*Aspergillus niger* Van Teighem), and leaf spots often occur together and cause pod yield loss up to 50–70% (Subrahmanyam et al. 1989; Mishra et al. 2015). Because of the frequent occurrences of multiple diseases, peanut yields are often significantly lower than their potential production (Holbrook and Stalker 2003).

Another challenge to enhancing peanut production is polyploidy ($2n = 4x = 40$) and a large genome size, which greatly complicates interpretation of genomic data as compared to the diploid wild relatives ($2n = 2x = 20$) (Guo et al. 2013). It is also difficult to transfer alleles from wild species to cultivated peanuts (Simpson 1991). During the past decade, extensive efforts in peanut genomics have resulted in a large number of genetic and genomic resources such as mapping populations, expressed sequence tags (ESTs), a wide range of molecular markers, transcriptomes, and proteomics analyses (Guo et al. 2013; Varshney et al. 2013; Katam et al. 2014), which were reviewed by Feng et al. (2012) and Guo et al. (2016). These genetic and genomic resources have been successfully used to construct genetic maps, to identify quantitative trait loci (QTL) of traits of interest, and to conduct marker-assisted selection and association mapping for peanut improvement (Pandey et al. 2014a, b, c; Guo et al. 2016).

Recognizing the challenges and importance of this crop and the benefits of enhancing our understanding of the peanut genome, the international peanut research community established the International Peanut Genomics Initiative (IPGI) in order to pool resources to meet these needs. In this chapter, we focus on the history and accomplishments of the IPGI in three areas: (1) brief background of the IPGI and a chronology of recent efforts in the peanut genome sequencing project; (2) recent developments in molecular markers, particularly molecular markers associated with disease resistance traits and current progress in marker-assisted breeding; and (3) recent efforts in developing next-generation populations for high-resolution genetic and trait mapping in peanut. Advances in each area over the years have come as the result of the initiative and international cooperative efforts of the scientific research community.

## 8.2 A Brief History and an Overview of the Peanut Genome Project

As early as 2001, the efforts were initiated at a meeting in Hunt Valley, MD on July 30–31, in light of the challenges and opportunities facing cool and warm season legume crops, and international research cooperation to develop new genomic technologies for legume crop improvement. Twenty-six legume scientists with knowledge of structural and functional genomics, DNA markers, transformation, bioinformatics, and legume crop improvement participated in a workshop hosted by the United Soybean Board, the National Peanut Foundation, the USA Dry Pea and Lentil Council, and the USDA-ARS to develop a strategy to advance genomics research across five economically important legume species. The group of scientists published the *U.S. Legume Crops Genomics White Paper* (Boerma et al. 2001) that outlined six areas where progress was needed across all legume crops. This meeting was followed by the Cross-Legume Advances through Genomics (CATG) Conference in Santa Fe, NM on December 14–15, 2004, where nearly 50 legume researchers and funding agency representatives met and developed a plan for cross-legume genomics research and to develop an action plan for legume research (Gepts et al. 2005). The peanut scientific community participated in both workshops. A book, *Legume Crop Genomics*, which documented the status of genomic resources for each legume crop including peanut (Wilson et al. 2004) was published under the auspices of the U.S. Legume Crop Genome Initiative (LCGI).

In 2004, 26 U.S. peanut scientists participated in a workshop hosted by the Peanut Foundation and American Peanut Council in Atlanta, GA on March 22–23. A *National Strategic Plan for the Peanut Genome Initiative* (PGI) (Wilson 2006b) was developed that outlined six objectives for the years 2004–2008: (1) improve the utility of genetic tools for peanut genomic research and develop useful molecular markers and genetic maps for peanut; (2) improve the efficacy of technology for gene manipulation in genomes and develop useful transformation methods for functional genomic research in peanut; (3) develop a framework for assembling the peanut genetic blueprint and locate abundant and rarely expressed genes, using genetic and physical approaches to integrate diverse data types; (4) improve knowledge of gene identification and regulation; (5) provide bioinformatics management of peanut biological information resources; and (6) determine the allergenic potential of peanut proteins. An action plan summarized in the white paper *National Program Action Plan for the Peanut Genome Initiative* soon followed (Wilson 2006a); and in 2006 an assessment of costs associated with genomic research was presented in the *Biotech Peanut White Paper "Benefits and Issues"* (Valentine et al. 2006) (http://www.peanutbioscience.com/images/PeanutWhitePaper.pdf).

In 2006, the PGI sought to expand its mission through outreach to the international peanut research community. The foundation for this effort was established in November 2006 in Guangzhou, China at the *International Conference on Aflatoxin Management and Genomics* where delegates from nine countries voted to maintain an open dialog to explore opportunities

for cooperative research, and to take steps toward achieving that goal with annual meetings. Therefore, the second conference of the international peanut research community was hosted on October 24–26, 2007 in Atlanta, GA, and it was the first conference to use the name Advances *in Arachis through Genomics & Biotechnology (AAGB): An International Strategic Planning Workshop* (credited to Varshney and ICRISAT team), which was another step toward bringing members of the international peanut community together to foster research collaboration on high priority issues. The *International Strategic Plan for the Peanut Genome Initiative 2008–2012 (IPGI): Improving Crop Productivity, Protection, and Product Safety & Quality* was developed at this workshop (http://www.peanutbioscience.com/images/Peanut_Genome_Intitative-StratPlan_DRAFT_v1.2_Mar08.pdf).

Since then the tradition of excellence that was established in Guangzhou and in Atlanta has been upheld at subsequent meetings including the *Third Advances in Arachis through Genomics & Biotechnology* (III AAGB-2008) at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) in India, IV AAGB-2009 in Bamako, Mali, V AAGB-2011 in Brasilia, Brazil, VI AAGB-2013 in Zhengzhou, China, VII AAGB-2014 in Savannah, Georgia, U.S., VIII AAGB-2015 in Brisbane, Australia, and IX AAGB-2017 in Cordoba, Argentina.

With the progress made by IPGI and the need to move on to the whole-genome sequencing discussion, the Peanut Foundation and the American Peanut Council on behalf of international peanut research community initiated the discussion of pursuing a peanut whole-genome sequencing project and related issues on July 12, 2010, at Clearwater, FL, as reported by Baozhu Guo, the liaison to China, that Chinese collaborators had been discussing a plan for peanut whole-genome sequencing. The executive committee of IPGI made the decision to send a delegation to China to initiate discussions with Chinese peanut collaborators for a possible joint sequencing project along with other international partners. Two members from the executive committee Victor Nwosu, Plant Science Program Manager, Global Chocolate Science & Technology of Mars Chocolate, NA, and Baozhu Guo, Plant Pathologist, USDA-ARS Crop Protection and Management Research Unit, traveled to China from September 2 to 12, 2010. The local host Xinyou Zhang, Henan Academy of Agricultural Sciences and Peanut Breeder, made the plan and arranged a meeting with Chairman Fuhe Luo who was then President of Guangdong Academy of Agricultural Sciences in 2006, and who had moved up to be national leader in Beijing since early 1998, along with other scientists, including Xingjun Wang of Shandong Academy of Agricultural Sciences and Da Luo of Sun Yat-Sen University. A proposal was made to collaborate and pool resources together to sequence both the tetraploid and diploid peanut genomes.

This trip report was sent to the IPGI executive committee, which started the ball rolling, resulting in the Peanut Genome Project Inaugural Meeting (http://www.peanutbioscience.com/peanutgenomeproject.html) on December 8 2010 in Atlanta, where Howard Shapiro of Mars also reported his meeting with BGI (Beijing Genomics Institute) concerning the peanut whole-genome sequencing. The Executive Committee of the IPGI agreed in principle to move the sequencing project forward to sequence peanut whole genomes with international collaborative effort. The IPGI Executive Committee called another meeting to continue the discussion of pursuing a peanut whole-genome sequencing project and related issues on January 12, 2011 at the Plant and Animal Genome (PAG) Conference, San Diego, CA, organized by Howard Shapiro and Rich Wilson, and tentatively decided to sequence four peanut cultivars (Tifrunner and GT-C20, SunOleic 97R, and NC94022) and their 200 recombinant inbred line (RIL) progenies (Qin et al. 2012) in collaboration with Chinese peanut researchers, in addition to the two diploid peanut progenitors. Nwosu and Guo made another trip to China from March 18 to 31, 2011 to discuss technical strategies and cost-sharing with Chinese collaborators, Xinyou Zhang, Suoyi Han, Wenyue Ma, Xingjun Wang,

Jiaquan Huang, Ronghua Tang, and Xuanqiang Liang along with Xiaoping Chen (on conference call). The group unanimously agreed to join the peanut genome sequencing projects and signed a Memorandum of Understanding.

A third trip was made by U.S. delegation to China (September 19–30, 2011) for meetings with Chinese collaborators and the China Ministry of Science and Technology, and discussed the time for launching the peanut genome sequencing project. The members were Victor Nwosu, Kim Moore, Howard Valentine, and Baozhu Guo. The sequencing and assembly strategies as proposed by BGI were adopting an integrated strategy combining whole-genome sequencing (WGS) plus bacterial artificial chromosomes (BACs) by BAC sequencing with Hiseq 2000 technology, and then resequencing and calling SNPs according to alignment to the developed reference genome with the aforementioned 200 RILs of the two mapping populations (Qin et al. 2012). The SNPs will then be used as markers to construct a genetic map for chromosome-level assembly.

In 2012, the sequencing and assembly strategies were discussed and adopted on March 28, in Atlanta. The Peanut Genome Consortium (PGC) was formally established as an extension of the IPGI and was embodied by a coalition of international scientists and stakeholders engaged in the Peanut Genome Project (PGP). PGP is an international collaborative research program whose goal is the complete mapping and understanding of all the genes of peanuts. PGC scientists have been deciphering the peanut genome in three major ways: developing polymorphic markers and producing genetic linkage maps; mapping the locations of genes/markers associated or linked with inherited traits such as disease resistance, yield, and quality; and determining the correct order or "sequence" of all the bases in peanut genome's DNA. Finally, the IPGI and PGC released the two diploid sequences for public use in April 2014 (http://www.peanutbase.org/node/618) and published the two genome sequences of the diploid ancestors of cultivated peanut in the journal *Nature Biotechnology* in February 2016 (Bertioli et al. 2016). The amazing findings of this publication were that these two genomes are very similar to the A and B subgenomes of allotetraploid cultivated peanut and could be used to identify candidate disease resistance genes, and to guide tetraploid transcript assemblies. Based on the high DNA identity of the *A. ipaensis* genome and the B subgenome of cultivated peanut and biogeographic evidence, the conclusion could be reached that *A. ipaensis* may be a direct descendant of the same population that contributed the B subgenome to cultivated peanut (Bertioli et al. 2016).

Another significant publication for *Arachis duranensis*, the peanut A-genome progenitor, a draft genome, was published in May 2016 in the journal of Proceedings of the National Academy of Sciences of the United States of America (PNAS) (Chen et al. 2016). This genome analysis suggests that the peanut lineage was affected by at least three polyploidizations since the origin of eudicots. Further resequencing of synthetic *Arachis* tetraploids also revealed extensive gene conversion since their formation by human hands. The *A. duranensis* genome provides a major source of candidate genes for fructification, oil biosynthesis, and allergens, expanding knowledge of understudied areas of plant biology and human health impacts of plants. This study also provides millions of structural variations that can be used as genetic markers for the development of improved peanut varieties through genomics-assisted breeding.

## 8.3 Major Contributions of IPGI in Trait Mapping and Molecular Breeding

During the years, much effort has been made to develop genetic and genomic tools and resources for cultivated peanut, such as construction of BAC libraries (Yuksel and Paterson 2005; Guimarães et al. 2008), cDNA libraries (Luo et al. 2005; Proite et al. 2007; Guo et al. 2008, 2009), RNAseq using next-generation sequencing technology (Guimaraes et al. 2012; Zhang et al. 2012), and development of DNA markers (see

reviews of Feng et al. 2012; Zhao et al. 2012; Varshney et al. 2013). Several reviews have recently been published in summarizing the achievements made in peanut genetics and genomics tool and resource development (see reviews of Feng et al. 2012; Zhao et al. 2012; Varshney et al. 2013; Guo et al. 2013, 2016).

Development of disease-resistant genotypes involves a series of processes including selection of ideal parents, generation of a large segregating population, and subsequent selection of desirable plants. Those are the essential steps of traditional interspecific hybridization breeding. Since 1960s, progress has been made in interspecific hybridization in peanuts because some wild *Arachis* species show a very high level of resistance to many diseases, such as ELS, LLS, rust, and stem rot (Holbrook and Stalker 2003). However, attempts to utilize these wild species as sources and the process of transferring the resistance and desired alleles to cultivated peanut has been severely hampered because of many factors, such as genomic (A and B genomes) and ploidy (diploid and tetraploid) barriers, restricted gene flow due to differences in ploidy level, the long period required for developing stable tetraploid interspecific derivatives, cross compatibility barriers, and a complicated inheritance mechanism (Burow et al. 2013). Meanwhile, conventional methods of screening germplasm in the field are time and resource consuming. Several factors contribute to the development of uniform occurrence of diseases in field conditions, which usually makes it difficult to achieve uniform distribution of disease pressure on populations and can lead to misclassification of lines (Yol et al. 2015). Moreover, the partial and polygenic nature of disease resistances makes the selection of ideal cross parents and the identification of resistant and susceptible lines in different generations very tedious using the traditional screening techniques (Leal-Bertioli et al. 2009).

Therefore, more efforts have been made to achieve progress in the area of crop genomics applied to breeding in recent years (Varshney et al. 2005; Guo et al. 2016). Combining genomics tools with conventional breeding can lead to more rapid development of resistant cultivars. Particularly, the advances in molecular marker technologies have provided techniques to improve crop breeding, which would be cost-effective and faster for selection, such as marker-assisted selection (MAS), which offers great promise for increasing the efficiency of conventional plant breeding, including the potential to pyramid resistance genes in peanut (Guo et al. 2012; Pandey et al. 2014a, b, c). Significant progress has been made for resistance to nematodes, rust, and leaf spots in addition to oil content and quality, and MAS has been applied into breeding programs for these traits (Simpson et al. 2003; Chu et al. 2011; Varshney et al. 2014; Khera et al. 2017).

### 8.3.1 Resistance to Root-Knot Nematode

Among the many pathogens of peanut, root-knot nematodes are among the most serious damaging and widespread (Dickson 1998). In many peanut production areas all over the world, root-knot nematodes are the most important factors that limit the yield of peanut and cause significant economic losses every year (Holbrook and Stalker 2003). There are three nematode species which can infect peanut, *Meloidogyne arenaria* (Neal) Chitwood, *M. hapla* Chitwood, and *M. javanica* (Treub) Chitwood. *Meloidogyne arenaria* is the predominant pathogenic species in the peanut areas of southern United States. About 40% fields in some areas can be infected resulting in yield losses in excess of 30% (Burow et al. 2014). Quantitative sources of resistance to root-knot nematodes have been identified in germplasm, even in *A. hypogaea*, and molecular work has been done to find the linked markers, genes, and QTLs.

Two dominant genes conferring resistance to *M. arenaria* were identified in an $F_2$ population, *Mae* and *Mag*. *Mae* is the gene restricting egg number, while *Mag* restricts galling. Meanwhile, a RAPD marker (Z3/265) which was linked to these genes was also identified (Garcia et al. 1996). These were the first molecular markers linked with a resistance gene for an

agronomically useful trait in peanut. Three more RAPD markers RKN410, RKN440, and KKN229 were found to be associated with nematode resistance in several backcross populations (Burow et al. 1996). These markers were all tightly linked and were all for the same gene. Two RFLP markers ca. 4 cM from the resistant gene were identified by bulked segregate analysis (Church et al. 2000). A sequence characterized amplified region (SCAR) marker, 197/909, was a new nematode resistance dominant marker. It was developed from the published sequence of a RAPD marker RKN440 and was found to be tightly linked with the resistance locus in populations derived from two tetraploid crosses (Burow et al. 1996; Chu et al. 2007). Two SSR markers with the genetic distance of 4.42 cM and 7.40 cM to root-knot nematode (*M. hapla*) resistance were discovered by analysis of an $F_2$ population derived from Huayu-22 and D099 (Wang et al. 2008). Nagy et al. (2010) developed a codominant SSR marker, GM565, through high-resolution mapping for nematode resistance, which could be used to identify heterozygotes for nematode resistance. These markers were then been examined for accuracy through field tests for root-knot nematode resistance in peanut (Branch et al. 2014). A new RFLP marker, R2430E, was found to be linked to the locus for the resistance to peanut root-knot nematode (*M. arenaria* race 1) (Pipolo et al. 2014).

Marker-assisted selection (MAS) has been demonstrated to be more efficient than phenotypic selection in use of markers for development of the nematode-resistant cultivars. Since Simpson et al. (1991) developed a root-knot nematode-resistant, synthetic allotetraploid line (TxAG-6), the first nematode-resistant peanut cultivar COAN was developed which contained a distinct trait donated from wild species (Simpson and Starr 2001). The second nematode-resistant peanut cultivar was then released with two generations of backcrossing of COAN-derived materials, and was developed by the use of RFLP markers (Simpson et al. 2003). These two resistant cultivars were found to carry the same dominant resistance gene for two root-knot nematode species (*M. arenaria* and *M. javanica*), and both have been widely used as important sources of resistance to root-knot nematode (Chu et al. 2007). For example, another nematode-resistant cultivar Tifguard was developed based on the improved nematode-resistant markers (Holbrook et al. 2008). This cultivar also was successfully converted into the Tifguard High O/L cultivar using MAS backcrossing selection (Chu et al. 2011). During the breeding process, three markers were involved including the dominant SCAR marker 197/199 (resistant allele), another dominant CAPS marker 1169/1170 (susceptible allele), and the codominant marker GM565 (Chu et al. 2011). This allowed for the identification of homozygous resistant, homozygous susceptible, and heterozygous individuals, respectively.

The previously mentioned markers for root-knot nematode resistance were mostly identified by using bulked segregate analysis. This method is efficient for identifying the markers with major effects but is less efficient for identifying the markers with minor effects (Burow et al. 2013). To improve sensitivity for small-effect QTLs, an advanced backcross population was screened for response to root-knot nematode infection. Composite interval mapping results suggested a total of seven QTLs plus three putative QTLs. These included the known major resistance gene plus the second QTL on LG1, and a potentially homeologous B-genome QTL on LG11. Additional potential homeologs were on LG8 and LG18, another QTL on LG9.2, and putative QTLs on LG9.1 and 19. Two introgressed QTLs were associated with susceptibility, and QTLs at some homeologous loci were found to confer opposite phenotypic responses (Burow et al. 2014).

### 8.3.2 Resistance to Leaf Spots and Rust

The foliar diseases early and late leaf spot are major destructive diseases of peanut worldwide (Shokes and Culbreath 1997). Epidemics of leaf spot diseases cause nearly complete defoliation

and yield losses of 50% or more through reduction of photosynthesis, death of the plant, and pod loss. Rust also is one of the severest diseases in peanut and can cause significant defoliation resulting in yield losses up to 50% (Subrahmanyam et al. 1989). Rust frequently occurs in combination with leaf spots, but one may predominate at different times (Burow et al. 2013). Although these diseases can be controlled by multiple applications of fungicides, long-term fungicide application could cause a slow erosion in disease control due to the gradual losses of sensitivity in the target population, environmental pollution, and economic impacts due to their application costs (Luo et al. 2005).

Because of the economic importance of these diseases, several studies in the area of molecular genetics and breeding have been performed, such as the application of different types of molecular markers, and the construction of peanut linkage maps (Mishra et al. 2015). Three RAPD markers associated with early leaf spot lesion diameter were identified through a population derived from the cross between an *A. cardenasii* introgression line and an *A. hypogaea* cultivated variety. Two breeding lines were developed from this genetic cross (Stalker and Mozingo 2001). A total of nine SSR markers were identified to be associated with rust resistance in two $F_2$ populations (Varma et al. 2005). Varman (1999) developed a rust-resistant line (VG9514) from the cross between the cultivar CO 1 and an *A. cardenasii* line to generate a mapping population. Using this population and a modified bulk segregate analysis, two RAPD markers, $J7_{1300}$ and $J7_{1350}$, linked to rust resistance were identified (Mondal et al. 2007). In the same lab, based on the Kruskal–Wallis one-way ANOVA and simple regression analysis, three and four SSR alleles were found associated with rust and late leaf spot, respectively (Mondal and Badigannavar 2010). Two genomic SSR markers (pPGPseq 4A05 and gi56931710) and a genic SSR marker (SSR-GO340445) for rust resistance in peanut were developed from the same mapping population mentioned previously (Mondal et al. 2012a, b). An $F_2$ population derived from Yuanza 9102 (a rust-susceptible line) and ICGV86699 (a rust-resistant cultivar) were applied to screen AFLP markers linked to rust resistance (Hou et al. 2007). By analysis of an $F_2$-segregating population derived from the cross of ICGV86699 and Zhonghua-5, three AFLP markers linked to late leaf spot resistance were identified (Xia et al. 2007). Using bulked segregate analysis, an SSR marker (PM384) was identified to have association with late leaf spot resistance. This marker could be utilized in marker-assisted breeding program (Shoba et al. 2012).

Burow et al. (2008) reported five markers for leaf spot resistance based on a mapping using RFLP markers, including three QTLs for incubation period and one each for latency period, lesion number, and diameter. Five QTLs were detected based on a mapping of 34 RGAs for late leaf spot disease resistance on detached leaves of the $F_2$ plants of the A-genome mapping population derived from *A. duranensis* × *A. stenosperma* (Leal-Bertioli et al. 2009), and suggested additive or partial dominance gene action. One QTL explained almost half of the phenotypic variance observed and two QTLs mapped near RGA markers. The first detailed study conducted in cultivated peanut was based on a partial genetic map comprising 56 SSR loci for the TAG24 × GPBD4 recombinant inbred line (Khedikar et al. 2010). This study reported 12 QTLs for rust (explaining 1.70–55.20% phenotypic variation). The SSR marker tightly linked to the major QTL (IPAHM103; QTLrust01) was then validated among a diverse set of genotypes as well as another population (Sarvamangala et al. 2011). This marker has been used for introgressing the major QTL for rust in peanut breeding program (Varshney et al. 2014). Using the same population mentioned previously and another RIL population, a consensus map with 225 SSR loci was developed. QTL analysis detected a total of 28 QTLs for late leaf spot and 15 QTLs for rust. A major QTL for late leaf spot ($QTL_{LLS}01$; linked markers GM1573 and Seq 8D09) with 10.27–62.34% phenotypic variance explained was detected across all the environments. Four new markers showed significant association with the major QTL (82.96% PVE) for rust resistance (Sujay et al. 2012). Validation of linked markers would accelerate the process of introgression of

rust and leaf spots resistance gene into preferred peanut genotypes. Gajjar et al. (2014) have attempted to validate the linkage of 22 SSR markers for rust and late leaf spot as reported by different workers, and 16 SSRs could be validated. QTL analysis based on an $F_2$ population derived from Tifrunner and GT-C20 had identified 37 QTLs for leaf spots, while in the $F_5$ map, 14 QTLs were found linked to leaf spots resistance (Wang et al. 2013). By using microarray analysis and real-time PCR, Luo et al. (2005) found genes were more greatly expressed in the resistant genotype as a response to *C. personatum* than in the susceptible genotype.

Another successful story of using MAS in peanut breeding was reported by Varshney et al. (2014) for rust resistance in addition to high oleic acid content and nematode resistance (Chu et al. 2011). Introgression of a major QTL for rust resistance through marker-assisted backcrossing has been successful in three popular Indian peanut cultivars, and generated several promising introgression lines with enhanced rust resistance and higher yield. One QTL explaining about 83% phenotypic variation for resistance to rust was validated and introgressed from the donor parent "GPBD 4" to three other peanut cultivars ("ICGV 91114", "JL 24", and "TAG 24") through marker-assisted breeding. There were a total of four markers used in the MAS breeding including one dominant (IPAHM103) and three codominant (GM2079, GM1536, GM2301) markers present in the QTL region (Varshney et al. 2014).

### 8.3.3 Resistance to Tomato Spotted Wilt Virus (TSWV)

Tomato spotted wilt virus is generally spread by thrips (*Frankinellia* spp.) and people usually control TSWV indirectly by applying insecticides. However, planting-resistant cultivars still is the best control strategy, which is effective and eco-friendly (Wang et al. 2013). Two major QTLs for TSWV resistance were identified for two RIL populations derived from the crosses of Tifrunner × GT-C20 and SunOleic 97R × NC94022, respectively, which explained 12.9% and 35.8% phenotypic variance (Qin et al. 2012). Recently, further study done in the different generations of the two populations identified 15 QTLs for TSWV resistance in $F_2$ map and nine QTLs in $F_5$ map, which explained 4.4–34.92% and 5.20–14.14% phenotypic variance, respectively (Wang et al. 2013). These were the only studies reporting QTL for TSWV resistance; however, it still provides hope for marker-assisted improvement of resistance to this disease (Burow et al. 2013). As a runner-type peanut cultivar, Tifrunner was released in 2005 with significantly higher resistance to TSWV than the moderately resistant cultivar Georgia Green (Holbrook and Culbreath 2007). Recently, Khera et al. (2017) reported an improved genetic linkage map for a recombinant inbred line (RIL) population derived from the cross between SunOleic 97R and NC94022. Multi-season (2010–2013) phenotypic data collected for the same population allowed for the identification of 16 major QTLs with more than 10% phenotypic variance explained, including four for resistance to TSWV, and six each for early spot and late leaf spot.

### 8.3.4 High Oleic Oil Peanuts

Oleic to linoleic acid ratios (O/L) in wild-type peanut are 1.0–4.0, whereas the O/L ratio in high oleic acid mutants is 35–40 (Norden et al. 1987). High O/L is desirable for healthy cholesterol-lowering benefits and the oxidative stability of the oil (Wilson et al. 2006a). The rate-limiting enzyme for the conversion of oleic to linoleic acid is oleoyl-PC desaturase (ahFAD2) (Ray et al. 1993). The two homoeologous genes encoding oleoyl-PC desaturase are *ahFAD2*A and *ahFAD2*B which are localized to the A and B subgenomes of *A. hypogaea*, respectively (Jung et al. 2000a, b). As an example of MAS in a breeding program for peanut cultivar improvement, an intensive backcross schedule has been developed to pyramid the high O/L trait with nematode resistance in the cultivar "Tifguard" (Holbrook et al. 2008). Crosses with

two high O/L cultivars, "Georgia 02C" (GAO2C) and "Florida 07", were made with "Tifguard". The markers used for nematode resistance were S197 and GM565 to detect the inheritance of the introgressed segment carrying *Rma*. Both high O/L donor parents possess the 441_442insA mutation which could be identified by CAPS marker *Hpy*188I, and all three parents carry the A-genome 448G → A transitional mutation in *ahFAD*2A allele; therefore, the only marker used for high O/L was *Hpy*188I-CAPS in order to track inheritance of high O/L. These markers can identify true hybrids at each stage of backcrossing. Therefore, the backcross and selection could be accelerated by using heterozygous $F_1$ hybrids as donor parents. In contrast to conventional breeding, which takes 8–10 years for a new cultivar release, this MAS approach is expected to produce a high O/L "Tifguard" within 26 months (Chu et al. 2011). Since then, efforts have been taken in ICRISAT and China and progress has been made significantly (Guo et al. 2016).

## 8.4 Recent Advancement in the Development of Next-Generation Mapping Populations for High-Resolution Genetic and Trait Mapping in Peanut

The primary goals of plant breeding for breeders are to improve yields, qualities, and other traits of commercial value suited to the needs of farmers and consumers (Moose and Mumm 2008). In practice, plant breeding mainly covers three processes: useful genetic variation is created or assembled; individuals with superior phenotypes are chosen; and improved cultivars are developed from selected individuals. The creation of experimental populations is a crucial step for plant breeders or geneticists (Varshney et al. 2006). However, during the initial period, breeder only depends on direct phenotypic selection, which is easily affected by genetic and environmental factors (Poormohammad Kiani et al. 2009). Thus, breeding methods depending only on phenotypic selection result in decreased accuracy and efficiency due to the fact that the majority of phenotypic variation in both natural populations and agricultural environments is determined by quantitative genetic traits (Mackay 2001).

With the advent of molecular marker, traditional breeding and genetics research is transitioning from a data-poor to a data-rich environment. Since the linkage mapping was developed using marker systems and crop traits (Edward et al. 1987; Paterson et al. 1988), the populations utilized for linkage mapping have included $F_2$, backcross (BC) or recombinant inbred (RI) populations and remain the primary methods used for plant QTL mapping studies (Huang et al. 2009). Recently, some researchers have applied genome-wide association study (GWAS) and single nucleotide polymorphism (SNP) genotyping markers into association analyses for crop genetic improvement (Rafalski 2010; Zhao et al. 2011; Riedelsheimer et al. 2012). Combined with next-generation sequencing technology, multi-parental mapping populations such as multi-parent advanced generation intercross (MAGIC) and nested association mapping (NAM) populations have become popular due to the high-resolution trait mapping obtained by combining the advantages of linkage analysis and association mapping (Zhang et al. 2005; Cavanagh et al. 2008; Yu et al. 2008). Recently, Huang et al. (2015) reviewed the current status and future prospects of multi-parent breeding populations, and here we summarize the efforts in the peanut community in developing next-generation multi-parental populations. We also compare the advantages and disadvantages of bi- and multi-parental mapping populations providing a frame of reference for choosing breeding populations in the future.

### 8.4.1 Bi-parental Breeding Populations

Bi-parental breeding populations, such as $F_2$, backcross (BC), and recombinant inbred line (RIL) populations, result from crosses between just two parents and may be effectively applied to

quantitative trait loci (QTL) analysis and linkage mapping (Verbyla et al. 2014). Among the bi-parental populations, the $F_2$ population is the easiest and most common population developed because it only requires a simple cross between two parents and the resulting progeny exhibits an obvious Mendelian segregation (Li et al. 2010). Genetic linkage map and QTL analyses with $F_2$ mapping populations have successfully been employed in peanut (Wang et al. 2013). However, the $F_2$ population still has two major limitations. First, the genetic structure of the $F_2$ population is easily affected by the environment resulting in difficulty of long-term preservation. Second, $F_2$ populations are the narrow genetic base of the population. For the dominant markers, the homozygous-dominant genotype and heterozygous genotypes could not be distinguished (Huang et al. 2015). To reduce these limitations and improve mapping resolution, recombinant inbred lines (RIL) populations are extensively utilized to map QTLs in peanuts (Qin et al. 2012; Pandey et al. 2014a, b, c; Wang et al. 2015; Guo et al. 2016). Compared to $F_2$ populations, RILs are permanent but cost time to create. With the development of next-generation sequencing technologies, RIL populations are easy to use in map construction and QTL mapping analysis for agronomic and morphological traits (Huang et al. 2009; Wang et al. 2010). Using peanut as an example, two major QTLs were detected in two related recombinant inbred line (RIL) populations (TAG 24 × GPBD 4 and TG 26 × GPBD 4) for rust resistance explaining up to 82.27% and 82.96% of phenotypic variance (PV), respectively (Sujay et al. 2012).

### 8.4.2 Multi-parent Advanced Generation Inter-Cross (MAGIC) Population

Multi-parent advanced generation intercross (MAGIC) populations provide an increased level of recombination and mapping resolution by integrating multiple alleles from different parents (Cavanagh et al. 2008). The MAGIC population can be developed using several different techniques. One method begins with a "funnel" breeding scheme also termed as a "classic MAGIC population", the multiple parents (founders) are intercrossed for n/2 generations (where "n" is the number of founders) until the founders are combined with equal proportions, followed by single seed descent (SSD) method to develop an RIL population (Rakshit et al. 2012). Another variant uses the half-diallele mating system for intercrossing the parents (two-way crosses) followed by intercrossing the $F_1$s until all the founders are represented in a single $F_1$ followed by SSD to the RIL population (Bandillo et al. 2013). In a simulation study in rice, Yamamoto et al. (2014) illustrated that the number of subsequent intercrosses dramatically increases the power of QTL detection. Recently, Verbyla et al. (2014) proposed a whole-genome average interval mapping (WGAIM) approach to simultaneously incorporate all founder probabilities at each marker for all individuals in the analysis, rather than using a genome scan in the R package "WGAIM", which could be useful in QTL analysis with multiple alleles.

The MAGIC scheme was first used in mice involving an eight-way cross using eight inbred strains and demonstrated that this population is efficient in fine mapping QTLs with small effects (Mott et al. 2000). Soon it was adapted in crops, and many populations in a wide range of species have been developed (Verbyla et al. 2014). Trebbi et al. (2008) developed an RIL population from a balanced four-way cross using four founders in durum wheat. In another study, two MAGIC populations were developed in wheat, one with four founders and the other with eight founders (Huang et al. 2012). In *Arabidopsis thaliana*, a MAGIC population containing 19 founders has been constructed (Kover et al. 2009). The most comprehensive MAGIC populations to date are in rice, where four MAGIC populations have been developed for the two subspecies *indica* and *japonica*. For the *indica* subspecies, *indica* MAGIC and MAGIC plus were developed containing eight *indica* parents. However, for *japonica* subspecies, *japonica* MAGIC and Global MAGIC have been developed containing eight *japonica* parents and 16

parents (eight *indica* and eight *japonica*), respectively (Bandillo et al. 2013). Development of a peanut MAGIC population with eight founders and 14 funnels is currently in progress (Huang et al. 2015) under the ambit of the Peanut Mycotoxin Innovation Lab (PMIL) project at ICRISAT, India. Another peanut MAGIC is also under development at Tifton, Georgia, USA with eight founders, which includes Tifrunner, GT-C20, SunOleic 97R, NC94022 (reported in Qin et al. 2012), Florida 07 and SPT-0606 (used in a peanut CAPS population as reported by Holbrook et al. 2013), and Georgia 13 M and TifNV-High O/L (two newly released cultivars) (Guo, personal communication).

### 8.4.3 Nested Association Mapping (NAM) Population

The nested association mapping (NAM) scheme is a proven strategy to dissect the genetic basis of complex traits in crops such as maize (Yu et al. 2008). The aim of the NAM design is to capture genetic diversity by selecting diverse parents (founders) and developing a large set of interrelated RIL mapping populations. An NGS platform then is used for generating dense genotyping data which helps in achieving high level of resolution by taking advantage of ancestral recombination. Because of this, the NAM population has higher QTL detection power as compared to individual bi-parental mapping populations (Yu et al. 2008; McMullen et al. 2009). In maize, the process of developing NAM populations involved individually crossing a set of 25 genetically diverse founders with a common parent "B73". The $F_1$s from each cross is forwarded through the SSD method to form an RIL population from each cross. The combined set of RILs arisen from each cross combination is called an NAM population (Yu et al. 2008). The 5000 lines developed from this effort in maize have been successfully implemented in dissecting several complex traits, such as flowering time (Buckler et al. 2009), 13 morphological traits (Brown et al. 2011), southern leaf blight resistance (Kump et al. 2011), northern leaf blight resistance (Tian et al. 2011), and kernel composition traits (Cook et al. 2012).

At ICRISAT, India, efforts are underway toward the development of NAM populations in peanut. In the U.S., the development of 16 structured RIL populations has been accomplished by crossing two common parental lines to eight unique lines (2 × 8) to generate two factorial nested association mapping populations (Holbrook et al. 2013). The common parents are Tifrunner and Florida-07 while the eight unique parents are N08082olJCT, C76-16, NC 3033, SPT 06-06, SSD 6, OLin, New Mexico Valencia A, and Florunner. These parents represent a wide range of disease resistance, agronomic, and morphological traits. Half of these RIL populations have been completed, and are being used by different research groups either individually to study unique traits, or as a whole by phenotyping and genotyping the RILs together as two mini NAM populations (total 1150 RILs) using Tifrunner and Florida-07 as two common parents and N08082olJCT, C76-16, NC 3033, SPT 06-06 as unique parents (Wang et al. 2016). This demonstrates its usefulness in assessing phenotypic diversity such as for morphological and disease resistance traits such as leaf length and width, plant size, main stem height, and leaf spot resistance which segregated within the assembled population and exhibit normal distributions. We also calculated the variance and heritability of each trait, and found that plant size had the lowest narrow sense heritability (0.06), while disease resistance had the highest (0.67) in the Tifrunner NAM population. In the Florida-07 population, main stem height had the lowest (0.27) and leaf width had the highest (0.73). Phenotyping of pod and kernel traits was very interesting and further genotyping by peanut SNP array is in progress. The NAM concept will promote the evaluation of the genetic diversity present in peanut gene pool.

## 8.5 Conclusion

From lack of sufficient molecular markers to the release of the genome sequences of two of its diploid wild relatives, international peanut

community has come a long way in the last 10 years. The international peanut genome project has been deciphering the peanut genome in three major ways: developing useful molecular markers and producing genetic maps; mapping QTLs and markers associated with important traits; and sequencing the whole peanut diploid and tetraploid genomes. The IPGI and PGC released the two diploid sequences for public use in April 2014 and published the two genome sequences in February 2016 (Bertioli et al. 2016). There is a long way to go before genomics-assisted breeding will be a routine tool for peanut improvement. Nevertheless, the stage is now set to harvest the fruits of genomics research, and it is expected that with the increasing effort toward SNP-based markers there will augment the use of GAB in peanut. It has been already proven that GAB is useful in developing high oleic, resistance to root-knot nematode, and rust resistance in peanut.

Additionally, the collaborative and coordinated efforts of the international peanut community since 2004 have contributed to the development of large-scale genomic resources and tools to tap into the rich resource of germplasm collections for improvement of peanut breeding for sustainable production, quality, pest resistance, and water use efficiency. With the establishment of NGS technology platforms and cost reduction for DNA sequencing, whole-genome sequencing and resequencing will become a routine task for crop research and improvement. The most challenging task will be the development of multi-parental populations and the integration of the new sequencing technology and the sequencing data being generated for tetraploid peanut for fine mapping and accurate trait identification and characterization. The main issue will be in analyzing data and translating the information to peanut breeding and improvement through the discovery of genes governing and molecular markers associated with the important traits.

**Acknowledgements** We would like to express our appreciation to the International Peanut Genome Initiative (IPGI) and Peanut Genome Consortium (PGC), U.S. Peanut Industries and the leadership of the Peanut Foundation and American Peanut Council, particularly to Howard Valentine for his efforts to make this IPGI and PGC and peanut genome sequencing project a reality. We also thank Dr. Howard Shapiro of Mars Inc. for initiating discussions with BGI and opening the way for collaborations with Chinese researchers. We are grateful for the financial support from USDA-ARS, the Peanut Foundation, the Georgia Peanut Commission and the U.S. National Peanut Board, Mars Inc., Bill & Melinda Gates Foundation (Tropical Legumes I, II & III), Department of Biotechnology (DBT) of Government of India, and World Bank Assisted Watershed Development Project II (KWDP-II) by Government of Karnataka, India; and genome collaborators from China (Henan Academy of Ag. Sci., Oil Crops Research Institute of Chinese Academy of Ag. Sci., and Hi-Tech Center of Shandong Academy of Ag. Sci.). The work reported in this article was undertaken as a part of the CGIAR Research Program on Grain Legumes. ICRISAT is a member of the CGIAR.

## References

Bandillo N, Raghavan C, Muyco PA et al (2013) Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice 6(1):11

Bertioli DJ, Cannon SB, Froenicke L et al (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet 48(4):438–446

Boerma HR et al (2001) U.S. legume crops genomics workshop White Paper. Hunt Valley, MD, 30–31 July 2001

Branch WD, Brenneman TB, Hookstra G (2014) Field test results versus marker assisted selection for root-knot nematode resistance in peanut. Peanut Sci 41(2):85–89

Brown PJ, Upadyayula N, Mahone GS et al (2011) Distinct genetic architectures for male and female inflorescence traits of maize. PLoS Genet 7(11): e1002383

Buckler ES, Holland JB, Bradbury PJ et al (2009) The genetic architecture of maize flowering time. Science 325(5941):714–718

Burow MD, Simpson CE, Paterson AH et al (1996) Identification of peanut (*Arachis hypogaea* L.) RAPD markers diagnostic of root-knot nematode (*Meloidogyne arenaria* (Neal) Chitwood) resistance. Mol Breeding 2(4):307–319

Burow MD, Starr JL, Park CH et al (2008) Identification of QTLs for resistance to early leaf spot (*Cercospora arachidicola* S. Hori) in an introgression population of peanut (*Arachis hypogaea* L.). In Proceedings of Plant and Animal Genome XVI, San Diego, p 424

Burow MD, Leal-Bertioli SCM, Simpson CE (2013) Marker-assisted selection for biotic stress resistance in

peanut. In:Varshney RK (ed) Translational Genomics for Crop Breeding, Biotic Stress, John Wiley & Sons Ltd: Chichester, UK, pp 125–150

Burow MD, Starr JL, Park CH et al (2014) Introgression of homeologous quantitative trait loci (QTLs) for resistance to the root-knot nematode [*Meloidogyne arenaria* (Neal) Chitwood] in an advanced backcross-QTL population of peanut (*Arachis hypogaea* L.). Mol breeding 34(2):393–406

Cavanagh C, Morell M, Mackay I et al (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Curr Opin Plant Biol 11(2):215–221

Chen X, Li H, Pandey MK et al (2016) Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens. Proc Natl Acad Sci 113:6785–6790

Chu Y, Holbrook CC, Timper P et al (2007) Development of a PCR-based molecular marker to select for nematode resistance in peanut. Crop Sci 47(2):841–847. doi:10.2135/cropsci2006.07.0474

Chu Y, Wu CL, Holbrook CC et al (2011) Marker-assisted selection to pyramid nematode resistance and the high oleic trait in peanut. Plant Genome 4(2):110–117. doi:10.3835/plantgenome2011.01.0001

Church GT, Simpson CE, Burow MD et al (2000) Use of RFLP markers for identification of individuals homozygous for resistance to *Meloidogyne arenaria* in peanut. Nematology 2(5):575–580

Cook JP, McMullen MD, Holland JB et al (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol 158(2):824–834

Dickson DW (1998) Peanut. Barker KR(ed) Plant and nematode interactions. American Society of Agronomy Inc, Madison, pp 523–566

Editorial Nature (2010) How to feed a hungry world. Nature 466:531–532

Edwards MD, Stuber CW, Wendel JF (1987) Molecular marker-facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116(1):113–125

Feng S, Wang X, Zhang X et al (2012) Peanut (*Arachis hypogaea*) expressed sequence tag (EST) project: progress and application. Comp Funct Genomics. doi:10.1155/2012/373768

Gajjar KN, Mishra GP, Radhakrishnan T et al (2014) Validation of SSR markers linked to the rust and late leaf spot diseases resistance in diverse peanut genotypes. Aust J Crop Sci 8(6):927–936

Garcia GM, Stalker HT, Shroeder E et al (1996) Identification of RAPD, SCAR, and RFLP markers tightly linked to nematode resistance genes introgressed from *Arachis cardenasii* into *Arachis hypogaea*. Genome 39(5):836–845

Gepts P, Beavis WD, Brummer EC et al (2005) Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through genomics conference. Plant Physiol 137(4):1228–1235

Guimaraes PM, Brasileiro ACM, Morgante CV et al (2012) Global transcriptome analysis of two wild relatives of peanut under drought and fungi infection. BMC Genom 13:387

Guimarães PM, Garsmeur O, Proite K et al (2008) BAC libraries construction from the ancestral diploid genomes of the allotetraploid cultivated peanut. BMC Plant Biol 8:14

Guo B, Chen X, Dang P et al (2008) Peanut gene expression profiling in developing seeds at different reproduction stages during *Aspergillus parasiticus* infection. BMC Dev Biol 8:12

Guo B, Chen X, Hong Y et al (2009) Analysis of gene expression profiles in leaf tissues of cultivated peanuts and development of EST-SSR markers and gene discovery. Int J Plant Genomics 2009:715605. doi:10.1155/2009/715605

Guo B, Chen CY, Chu Y et al (2012) Advances in genetics and genomics for sustainable peanut production. In: Benkeblia N (ed) Sustainable Agriculture and New Biotechnologies. CRC Press, Boca Raton, FL pp 341–367

Guo B, Pandey MK, He G et al (2013) Recent advances in molecular genetic linkage maps of cultivated peanut (*Arachis hypogaea* L.). Peanut Sci 40(2):95–106. doi:10.3146/PS13-03.1

Guo B, Khera P, Wang H et al (2016) Annotation of trait loci on integrated genetic maps of *Arachis* species. In: Stalker T and Wilson R (eds) Peanuts: Genetics, processing, and utilization. The American Oil Chemists' Society (AOCS), San Diego, pp 163–208

Holbrook CC, Culbreath AK (2007) Registration of 'Tifrunner'peanut. J Plant Reg 1(2):124

Holbrook CC, Stalker HT (2003) Peanut breeding and genetic resources. Plant Breed Rev 22:297–356

Holbrook CC, Timper P, Culbreath A et al (2008) Registration of 'Tifguard' peanut. J Plant Reg 2 (2):92–94

Holbrook CC, Isleib TG, Ozias-Akins P et al (2013) Development and phenotyping of recombinant inbred line (RIL) populations for peanut (*Arachis hypogaea*). Peanut Sci 40(2):89–94

Hou H, Liao B, Lei Y et al (2007) Identification of AFLP markers for resistance to peanut rust. Chin J Oil Crop Sci 29(2):89–92

Huang X, Feng Q, Qian Q et al (2009) High-throughput genotyping by whole genome resequencing. Genome Res 19:1068–1076

Huang BE, George AW, Forrest KL et al (2012) A multiparent advanced generation inter-cross population for genetic analysis in wheat. Plant Biotechnol J 10(7):826–839

Huang BE, Verbyla KL, Verbyla AP et al (2015) MAGIC populations in crops: current status and future prospects. Theor Appl Genet 128(6):999–1017

Jung S, Powell G, Moore K et al (2000a) The high oleate trait in the cultivated peanut [*Arachis hypogaea* L.]. II. Molecular basis and genetics of the trait. Mol Gen Genet 263(5):806–811

Jung S, Swift D, Sengoku E et al (2000b) The high oleate trait in the cultivated peanut [*Arachis hypogaea* L.]. I. Isolation and characterization of two genes encoding microsomal oleoyl-PC desaturases. Mol Gen Genet 263(5):796–805

Katam R, Gottschalk V, Survajahala P et al (2014) Advances in proteomics research for peanut genetics and breeding. In: Mallikarjuna N, Varshney RK (eds) Genetics, Genomics and Breeding of Peanuts. CRC Press, Boca Raton, pp 161–177

Khedikar YP, Gowda MVC, Sarvamangala C et al (2010) A QTL study on late leaf spot and rust revealed one major QTL for molecular breeding for rust resistance in groundnut (*Arachis hypogaea* L.). Theor Appl Genet 121(5):971–984

Khera P, Pandey MK, Wang H, Feng S, Qiao L, Culbreath AK, Kale S, Wang J, Holbrook CC, Zhuang W, Varshney RK, Guo BZ (2017) Mapping quantitative trait loci of resistance to Tomato spotted wilt virus and leaf spots in a recombinant inbred line population of peanut (*Arachis hypogaea L.*) from SunOleic 97R and NC94022. PLoS ONE (accepted)

Kover PX, Valdar W, Trakalo J et al (2009) A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. PLoS Genet 5(7): e1000551

Kump KL, Bradbury PJ, Wisser RJ et al (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat Genet 43 (2):163–168

Lander ES, Linton LM, Birren B et al (2001) Initial sequencing and analysis of the human genome. Nature 409(6822):860–921

Leal-Bertioli SC, José AC, Alves-Freitas DM et al (2009) Identification of candidate genome regions controlling disease resistance in *Arachis*. BMC Plant Biol 9:112

Li Y, Willer CJ, Ding J et al (2010) Using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 34(8):816–834

Luo M, Dang P, Guo B et al (2005) Generation of expressed sequence tags (ESTs) for gene discovery and marker development in cultivated peanut. Crop Sci 45(1):346–353

Mackay TFC (2001) The genetic architecture of quantitative traits. Ann Rev Genet 35:303–339

McMullen MD, Kresovich S, Villeda HS et al (2009) Genetic properties of the maize nested association mapping population. Science 325(5941):737–740

Mishra GP, Radhakrishnan T, Kumar A et al (2015) Advancements in molecular marker development and their applications in the management of biotic stresses in peanuts. Crop Prot 77:74–86

Mondal S, Badigannavar AM (2010) Molecular diversity and association of SSR markers to rust and late leaf spot resistance in cultivated groundnut (*Arachis hypogaea* L.). Plant Breeding 129(1):68–71

Mondal S, Badigannavar AM, Murty GSS (2007) RAPD markers linked to a rust resistance gene in cultivated groundnut (*Arachis hypogaea* L.). Euphytica 159 (1):233–239

Mondal S, Badigannavar AM, D'Souza SF (2012a) Development of genic molecular markers linked to a rust resistance gene in cultivated groundnut (*Arachis hypogaea* L.). Euphytica 188(2):163–173

Mondal S, Badigannavar AM, D'Souza SF (2012b) Molecular tagging of a rust resistance gene in cultivated groundnut (*Arachis hypogaea* L.) introgressed from *Arachis cardenasii*. Mol Breed 29(2): 467–476

Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiol 147(3):969–977

Mott R, Talbot CJ, Turri MG et al (2000) A method for fine mapping quantitative trait loci in outbred animal stocks. Proc Natl Acad Sci 97(23):12649–12654

Nagy E, Chu Y, Guo Y et al (2010) Recombination is suppressed in an alien introgression in peanut harboring *Rma*, a dominant root-knot nematode resistance gene. Mol Breed 26(2):357–370

Norden AJ, Corbet DW, Knauft DA et al (1987) Variabilityin oil quality among peanut genotypes in the Florida breeding program. Peanut Sci 14(1):7–11

Pandey MK, Guo B, Holbrook CC et al (2014a) Molecular markers, genetic maps and QTLs for molecular breeding in peanut. In: Mallikarjuna N (ed) Genetics, genomics and breeding of peanuts. CRC Press, Boca Raton, pp 61–113

Pandey MK, Wang ML, Qiao L et al (2014b) Identification of QTLs associated with oil content and mapping *FAD2* genes and their relative contribution to oil quality in peanut (*Arachis hypogaea* L.). BMC Genet 15:133

Pandey MK, Upadhyaya HD, Rathore A et al (2014c) Genome wide association studies for 50 agronomic traits in peanut using the 'Reference Set' comprising 300 genotypes from 48 countries of the semi-arid tropics of the world. PLoS ONE 9(8):e105228

Paterson AH, Lander ES, Hewitt JD et al (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335(6192):721–726

Pípolo VC, Meagher MG, Dickson DW et al (2014) Molecular marker screening of peanut (*Arachis hypogaea* L.) germplasm for *Meloidogyne arenaria* resistance. Afr J Biotechnol 13(26):2608–2612

Poormohammad Kiani S, Maury P, Nouri L et al (2009) QTL analysis of yield-related traits in sunflower under different water treatments. Plant Breeding 128(4):363–373

Proite K, Leal-Bertioli SC, Bertioli DJ et al (2007) ESTs from a wild *Arachis* species for gene discovery and marker development. BMC Plant Biol 7:7

Qin H, Feng S, Chen C et al (2012) An integrated genetic linkage map of cultivated peanut (*Arachis hypogaea* L.) constructed from two RIL populations. Theor Appl Genet 124(4):653–664

Rafalski JA (2010) Association genetics in crop improvement. Curr Opin Plant Biol 13(2):174–180

Rakshit S, Rakshit A, Patil JV (2012) Multiparent intercross populations in analysis of quantitative traits. J Genet 91(1):111–117

Ray TK, Holly SP, Knauft DA et al (1993) The primary defect in developing seed from the high oleate variety of peanut (*Arachis hypogaea* L.) is the absence of $\Delta^{12}$–desaturase activity. Plant Sci 91(1):15–21

Riedelsheimer C, Lisec J, Czedik-Eysenberg A et al (2012) Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. Proc Natl Acad Sci 109(23):8872–8877

Sarvamangala C, Gowda MVC, Varshney RK (2011) Identification of quantitative trait loci for protein content, oil content and oil quality for groundnut (*Arachis hypogaea* L.). Field Crops Res 122(1):49–59

Shoba D, Manivannan N, Vindhiyavarman P et al (2012) SSR markers associated for late leaf spot disease resistance by bulked segregant analysis in groundnut (*Arachis hypogaea* L.). Euphytica 188(2):265–272

Shokes FM, Culbreath AK (1997) Early and late leaf spots. In: Kokalis-Burrell N (ed) Compendium of peanut diseases. APS Press, Paul, pp p17–p20

Simpson CE (1991) Pathways for introgression of pest resistance into *Arachis hypogaea* L. Peanut Sci 18 (1):22–26

Simpson CE, Starr JL (2001) Registration of 'COAN' peanut. Crop Sci 41(3):918

Simpson CE, Starr JL, Church GT et al (2003) Registration of NemaTAM peanut. Crop Sci 43(4):1561

Stalker HT, Mozingo IG (2001) Molecular markers of *Arachis* and marker-assisted selection. Peanut Sci 28 (2):117–123

Subrahmanyam P, Ramanatha Rao V, McDonald D et al (1989) Origins of resistances to rust and late leaf spot in peanut (*Arachis hypogaea*, Fabaceae). Econ Bot 43 (4):444–455

Subramanian G, Adams MD, Venter JC et al (2001) Implications of the human genome for understanding human biology and medicine. J Am Med Assoc 286 (18):2296–2307

Sujay V, Gowda MVC, Pandey MK et al (2012) QTL analysis and construction of consensus genetic map for foliar diseases resistance based on two RIL populations in cultivated groundnut (*Arachis hypogaea* L.). Mol Breed 30(2): 773–788

Tian F, Bradbury PJ, Brown PJ et al (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet 43 (2):159–162

Trebbi D, Maccaferri M, Giuliani S et al (2008) Development of a multi-parental (four-way cross) mapping population for multiallelic QTL analysis in durum Wheat. In: Appels R, Eastwood R, Lagudah E, Langridge P, Mackay M, McIntyre L, Sharp P (eds.) Proceedings of the 11th international wheat genetics symposium, Sydney University, Australia, 24–29 August 2008

Valentine H, Grabau E, Ozias-Akins P et al (2006) Biotech Peanut White Paper: Benefits and Issues. VA, American Peanut Council, Alexandria, p 14

Varma TSN, Dwivedi SL, Pande S et al (2005) SSR markers associated with resistance to rust (Puccinia arachidis Speg.) in groundnut (*Arachis hypogaea* L.). SABRAO J Breed Genet 37(2):107–119

Varman PV (1999) A foliar disease resistant line developed through interspecific hybridization in groundnut (*Arachis hypogaea*). Indian J Agri Sci 69(1):67–68

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23(1):48–55

Varshney RK, Hoisington DA, Tyagi AK (2006) Advances in cereal genomics and applications in crop breeding. Trends Biotechnol 24(11):490–499

Varshney RK, Mohan SM, Gaur PM et al (2013) Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. Biotechnol Adv 31(8):1120–1134

Varshney RK, Pandey MK, Janila P et al (2014) Marker-assisted introgression of a QTL region to improve rust resistance in three elite and popular varieties of peanut (*Arachis hypogaea* L.). Theor Appl Genet 127(8):1771–1781

Venter JC, Adams MD, Myers EW et al (2001) The sequence of the human genome. Science 291 (5507):1304–1351

Verbyla AP, George AW, Cavanagh CC et al (2014) Whole genome QTL analysis for MAGIC. Theor Appl Genet 127(8):1753–1770

Wang H, Shi Y, Ren Y et al (2008) Development of SSR markers for root-knot nematode resistance in peanut. Journal of Peanut Science 37(2):14–17

Wang Q, Zhang X, Tang F et al (2010) Construction of genetic linkage map of peanut (*Arachis hypogaea* L.) based on SRAP markers. Chin J Oil. Crop Sci 32 (3):374–378

Wang H, Pandey MK, Qiao L et al (2013) Genetic mapping and quantitative trait loci analysis for disease resistance using F2 and F5 generation-based genetic maps derived from 'Tifrunner' × 'GT-C20' in peanut. Plant Genome 6(3):1–10

Wang ML, Khera P, Pandey MK et al (2015) Genetic mapping of QTLs controlling fatty acids provided insights into the genetic control of fatty acid synthesis pathway in peanut (*Arachis hypogaea* L.). PLoS ONE 10(4):e0119454

Wang H, Guo X, Pandey M et al (2016) Phenotypic assessments of peanut nested association mapping (NAM) populations. Phytopathology 106(S2):15

Wilson RF (2006a) National Program Action Plan for the Peanut Genome Initiative: Application of Plant Genomics to Mitigate Peanut Allergy. Beltsville, USDA, p 25

Wilson RF (2006b) National Strategic Plan for the Peanut Genome Initiative 2004–2008. Beltsville, USDA, p 19

Wilson RF, Stalker HT, Brummer CE (2004) Genomics for Legume Crops. Amer Oil Chemical Soc Press, Champaign, p 362

Xia Y, Liao B, Li J et al (2007) Identification of AFLP markers linked to resistance to late leaf spot in peanut

(*Arachis hypogaea* L.). Chin J Oil. Crop Sci 29 (3):318–321

Yamamoto E, Iwata H, Tanabata T et al (2014) Effect of advanced intercrossing on genome structure and on the power to detect linked quantitative trait loci in a multi-parent population: a simulation study in rice. BMC Genet 15:50

Yol E, Upadhyaya HD, Uzun B (2015) Molecular diagnosis to identify new sources of resistance to sclerotinia blight in groundnut (*Arachis hypogaea* L.). Euphytica 203(2):367–374

Yu J, Holland JB, McMullen MD et al (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178(1):539–551

Yuksel B, Paterson AH (2005) Construction and characterization of a peanut HindIII BAC library. Theor Appl Genet 111(4):630–639

Zhang M, Montooth KL, Wellset MT et al (2005) Mapping Multiple Quantitative Trait Loci by Bayesian Classification. Genetics 169(4):2305–2318

Zhang J, Liang S, Duan J et al (2012) De novo assembly and Characterisation of the Transcriptome during seed development, and generation of genic-SSR markers in Peanut (*Arachis hypogaea* L.). BMC Genom 13:90

Zhao K, Tung CW, Eizenga GC et al (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat Commun 2:467

Zhao Y, Prakash CS, He G (2012) Characterization and compilation of polymorphic simple sequence repeat (SSR) markers of peanut from public database. BMC Res Notes 5:362

# 9 Sequencing Ancestor Diploid Genomes for Enhanced Genome Understanding and Peanut Improvement

Spurthi N. Nayak, Manish K. Pandey, Scott A. Jackson, Xuanqiang Liang and Rajeev K. Varshney

**Abstract**

Cultivated peanut (*Arachis hypogaea*) is an allotetraploid with closely related subgenomes of a total size of ~2.7 Gb. To understand the genome of the cultivated peanut, it is prerequisite to know the genome organization of its diploid progenitors, A-genome—*Arachis duranensis* and B-genome—*A. ipaensis*. Two genome sequencing projects conducted sequencing and analysis of the genomes of diploid ancestors: (1) International Peanut Genome Initiative (IPGI) reported the sequencing of both A- and B-genomes; while (2) Diploid Progenitor Peanut Arachis Genome Sequencing Consortium (DPPAGSC) reported the sequencing of A-genome. IPGI study showed that these genomes are similar to cultivated peanut's A- and B-subgenomes and used them to identify candidate disease resistance genes, to guide tetraploid transcript assemblies and to detect genetic exchange between cultivated peanut's subgenomes thus providing evidence about direct descendant of the B subgenome in cultivated peanut. The DPPAGSC study, on the other hand, provided new insights into geocarpy, oil biosynthesis, and allergens in addition to providing information about evolution and polyploidization. These genome sequencing efforts have improved the understanding about the complex peanut genome and genome architecture which will play a very important role in peanut applied genomics and breeding.

S.N. Nayak · M.K. Pandey · R.K. Varshney (✉)
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India
e-mail: r.k.varshney@cgiar.org

S.A. Jackson
Center for Applied Genetic Technologies, University of Georgia (UGA), Athens, USA

X. Liang
Crop Research Institute, Guangdong Academy of Agricultural Sciences (GAAS), Guangzhou, China

R.K. Varshney
University of Western Australia (UWA), Crawley, Australia


R.K. Varshney et al. (eds.), *The Peanut Genome*,
Compendium of Plant Genomes, DOI 10.1007/978-3-319-63935-2_9

## 9.1 Introduction

Since the availability of first plant genome *Arabidopsis thaliana* in 2000, genomes of several plant species have been sequenced (Michael and Jackson 2013). With advancements in sequencing technologies and genome assembly methodologies over the past decade, genome sequencing is now not limited to only model plant species or small genomes. Several crop plants, plantation crops, vegetables, fruits and even the wild progenitors of important crop species have been sequenced and many are in progress. With the advent of next-generation sequencing (NGS) technologies, there is a rapid increase in sequenced plant genomes due to the exponential decrease in cost and time in generating sequencing data (Varshney et al. 2009; Schatz et al. 2012). Rice was the first sequenced crop genome and had a major impact on accelerating rice genetics research and breeding applications (Jackson 2016). The genome sequencing projects for most crops have been possible due to international collaborations and both formal and informal consortia.

Most of the sequenced plants have been diploids while the sequencing of polyploids and large sized genomes has been less frequent. Polyploid genomes increase the genome complexity and therefore, pose a serious challenge towards the development of high-quality assemblies of pseudomolecules and genomes. Hence as a basis for the polyploid genome sequencing, where available, the diploid progenitors have been sequenced for several polyploid plant species like cotton (Wang et al. 2012), wheat (Ling et al. 2013, Jia et al. 2013, Marcussen et al. 2014), and capsicum (Qin et al. 2014). Polyploidy or whole-genome duplication (WGD) has been proposed to be a major evolutionary force in plants, especially in angiosperms (see Soltis et al. 2014). Cultivated peanut is an allotetraploid with total genome size of ~2.7 Gb. The peanut subgenomes are closely related (Nielen et al. 2012; Moretzsohn et al. 2013). However, the A and B subgenomes appear to have undergone relatively few changes since polyploidization as evidenced by genomic in situ hybridization (GISH) which clearly distinguished A and B chromosomes without much mosaics (Ramos et al. 2006; Seijo et al. 2007). The genome size of *A. hypogaea* is close to the sum of those for *A. duranensis* (1.25 Gb) and *A. ipaensis* (1.56 Gb), indicating that there has been no large change in genome size since polyploidy (Samoluk et al. 2015). In addition, progenies derived from crosses between cultivated peanut and an artificially induced allotetraploid (*A. ipaensis* K30076 × *A. duranensis* V14167) ($2n = 4x = 40$) were fertile and phenotypically normal with low segregation distortion (Foncéka et al. 2009). These observations strongly support the close relationships between the diploid genomes of the progenitors and the corresponding subgenomes of *A. hypogaea* (Fávero et al. 2006). Hence sequencing of diploid progenitors was a logical choice as it not only provides ease of tetraploid assembly, but also provides a deeper understanding of *Arachis* biology, evolution and any genomic change following polyploid formation.

## 9.2 Sequencing of Progenitor Diploid Genomes of Cultivated Peanut

Sequencing of the peanut A-genome progenitor, *A. duranensis* V14167, and the B-genome progenitor, *A. ipaensis* K30076, was completed by the International Peanut Genome Initiative (IPGI, http://www.peanutbioscience.com/peanutgenomeinitiative.html) and published in *Nature Genetics* (Bertioli et al. 2016) (Table 9.1). In another effort, the A-genome progenitor, *A. duranensis* PI475845 was sequenced by China-ICRISAT-UGA co-led initiative (Diploid Progenitor Peanut A-Genome Sequencing Consortium, DPPAGSC, http://ceg.icrisat.org/dppga/Manuscript.html) and published in *Proceedings of National Academy of Sciences of the United States of America* (Chen et al. 2016) (Table 9.1). The genotype V14167 (A-genome, *A. duranensis*) originated from Argentina while the other two genotypes, PI 475845 (A-genome, *A. duranensis*) and K30076 (B-genome, *A. ipaensis*), originated in Bolivia.

## 9.3 Strategies and Tools for Sequencing

### 9.3.1 Sequencing Platform

The Illumina HiSeq 2000/2500 platforms were used to generate sequence data in the peanut genome projects. Illumina captures template DNA that has been ligated to specific adapters in a flow cell, a glass enclosure similar in size to a microscope slide, with a dense lawn of primers. The template is then amplified into clusters of identical molecules, or polonies, and sequenced in cycles using DNA polymerase. Terminator dNTPs in the reaction are labeled with different fluorescent labels and detection is by optical fluorescence. As only terminators are used, only one base can be incorporated in one cluster in every cycle. After the reaction is imaged in four different fluorescence levels, the dye and terminator group is cleaved off and another round of dye-labeled terminators is added. The total number of cycles determine the length of the read. While generating peanut genome sequences, the read lengths ranged from 90–150 bp.

### 9.3.2 Sequence Data Generation

The sequence data were generated using paired-end sequencing insert libraries with insert sizes of 250 bp, 500 bp, 2, 5, 10 and 20 kb using standard protocols provided by Illumina (San Diego, USA). The sequencing yielded in 325.73 Gb of raw data reads for *A. duranensis* and 416.59 Gb for *A. ipaensis* under the IPGI project whereas 229.94 Gb raw data was obtained from *A. duranensis* for the DPPAGSC project.

### 9.3.3 Quality Filtering

Reads with more than 5% Ns or with polyadenylated termini; reads from the short-insert libraries (170–800 bp) with 20 or more bases having quality scores $\leq 7$; reads from the large-insert libraries (2–40 kb) with 40 or more bases having quality score $\leq 7$; reads with adaptor contamination (more than 10 bp aligned to the adaptor sequence when allowing $\leq 3$ bp of mismatches); reads with read 1 and read 2 having $\geq 10$ bp overlapping (allowing 10% mismatches; except for the 250-bp insert library, where the paired reads should overlap); reads identical to each other at both ends that might have been caused by PCR duplication; and reads where the quality of the bases at the head or tail was $\leq 7$ were discarded in US-led initiative.

Under DPPAGSC project, the reads of short-insert libraries were trimmed of four low-quality bases at both ends, and reads of long-insert libraries were trimmed of three low-quality bases; duplicated reads from long-insert libraries were filtered out; the reads with 10 or more Ns (no sequenced bases) and low-quality bases were also filtered out from individual reads in all lanes.

### 9.3.4 *k*-mer Analysis

*k*-mers were extracted from sequences generated from the short-insert libraries, and the frequencies were calculated and plotted. Genome sizes were estimated by dividing the total numbers of *k*-mers by the depths of the major peaks.

**Table 9.1** Summary of genome sequencing efforts for diploid progenitor species

| Progenitor species | Genome | Genotype sequenced | Assembly size (Gb) | Genes predicted | Lead consortium |
|---|---|---|---|---|---|
| *A. duranensis* | A | V14167 | 1.21 | 36,734 | USA-led IPGI |
| *A. duranensis* | A | PI475845 | 1.05 | 50,324 | China-led DPPAGSC |
| *A. ipaensis* | B | K30076 | 1.51 | 41,840 | USA-led IPGI |

### 9.3.5 Error Correction

*k*-mers were used to correct for errors. For sequencing with high depth, the *k*-mers without any sequencing errors should appear multiple times in the read data set, whereas error-containing *k*-mers should have low frequencies. Sequencing errors in the 17-mers with frequencies lower than three in the clean data for the 250- and 500-bp insert libraries were corrected.

## 9.4 Tools and Technology Used in Genome Assembly

Under the IPGI project, COPE (Liu et al. 2012) was used to join paired-end reads from the 250-bp insert library into single longer reads of ~250 bp. Genome assembly was performed using SOAPdenovo version 2.05 (Li et al. 2010), with parameters –K 81 –R. Gaps were filled using KGF and Gapcloser version 1.10 (Luo et al. 2012). Finally, SSPACE (Boetzer et al. 2011) was used to further link the scaffolds where connections were supported by more than five paired reads. For assembling genome under DPPAGSC project, 159.07Gb filtered reads were further used for genome assembling. SOAPdenovo2 (version 2.04.4) with optimized parameters (pregraph -K 79 -p 16 -d 5; scaff -F -b 1.5) was used to construct contigs and original scaffolds. Newbler and SOAPdenovo were used with parameters -K 79 -p 16 -d 5. The gaps were closed with GapCloser, scaffolds were reconstructed using Haplomerger (Huang et al. 2012). The paired-end information was subsequently applied to link contigs into scaffolds in a stepwise manner. Several intra-scaffold gaps were filled by local assembly using the reads in a read-pair, where one end uniquely mapped to a contig, whereas the other end was located within a gap. Subsequently, SSPACE (version 2.0; using core parameters "-k 6 -T 4 -g 2") was used to link the SOAPdenovo2 scaffolds.

Under IPGI project, ultradense genetic maps were generated through genotyping-by-sequencing (GBS) of two diploid recombinant inbred line (RIL) populations. SNPs within scaffolds were used to validate the assemblies and confirmed their high quality. Based on the presence of diagnostic population-wide switches in SNP genotypic data occurring at the point of misjoin, 190 of 1297 initial scaffolds of *A. duranensis* and 49 of 353 initial scaffolds of *A. ipaensis* were identified as chimeric. These chimeric scaffolds were split and used for remapping. Thus, approximate chromosomal placements were obtained for 1692 and 459 genetically verified scaffolds, respectively. Conventional linkage maps along with the syntenic inferences were used to refine the ordering of scaffolds within the initial genetic bins. Generally, agreement was good for maps in euchromatic arms and poorer in pericentromeric regions. Overall, 96.0 and 99.2% of the sequence in contigs $\geq$ 10,000 bp in length, represented by 1692 and 459 scaffolds, could be ordered into 10 chromosomal pseudomolecules per genome of 1025 and 1338 Mb for *A. duranensis* and *A. ipaensis*, respectively. The pseudomolecules were named as Aradu.A01–Aradu.A10 (GCA_000817695.1) and Araip.B01–Araip.B10 (GCA_000816755.1). The pseudomolecules mostly showed one-to-one equivalence between the A- and B-genomes and were numbered according to previously published linkage maps (Shirasawa et al. 2013, Gautami et al. 2012, Moretzsohn et al. 2005, 2009). They represent 82% and 86% of the genomes, respectively, when considering genome size estimates based on flow cytometry, or 95 and 98% of the genomes when using estimates derived from *k*-mer frequencies with $k = 17$. Comparisons of the chromosomal pseudomolecules with 14 BAC sequences from *A. duranensis* and 6 BAC sequences from *A. ipaensis* showed collinearity of contigs and high sequence identity ($\geq$ 99%). This information was used to improve the genome assembly to pseudomolecule level under IPGI whereas, DPPAGSC has assembly that contained 8173 scaffolds.

### 9.4.1 Production of Moleculo Synthetic Long Reads

In IPGI project, the TruSeq synthetic long-read sequencing libraries (McCoy et al 2014) were

generated by Moleculo and Illumina as part of beta tests of this technology. Fifteen libraries were generated for *A. duranensis* K7988, and each library was sequenced on a HiSeq 2500 lane; the PE100 reads were assembled into 1.5 million TruSeq (Moleculo) synthetic long reads, providing approximately 5X genome coverage with a mean read length of 3684 bases and an N50 of 4344 bases. Twelve libraries were used for *A. ipaensis* K30076 to yield approximately 2 million Moleculo reads with mean length of 4054 bases and an N50 length of 5152 bases, providing ~6X genome coverage. Thirteen libraries were used for *A. hypogaea* cv. Tifrunner, which produced 1263,111 Moleculo reads with a mean length of 4547 bases and an N50 length of 6137 bases, providing 2.3X genome coverage. These reads were used for genome comparisons and were not incorporated in the diploid genome assemblies.

### 9.4.2 Linkage Maps and Identification of Misjoins

Conventional molecular marker maps from diploid A- and B-genomes and cultivated peanut × induced allotetraploid recombinant inbred lines (RIL) populations were used to find the order of the scaffolds from peanut assembly. Genetic maps generated from genotyping-by-sequencing data for diploid A- and B-genome RIL populations were used in identification of chimeric scaffolds. RILs from the diploid A- and B-genome populations were shotgun sequenced to 1X genome coverage with paired-end 100-bp reads on a HiSeq 2500 sequencer. The parents were sequenced at 20X genome coverage. Parental-homozygous SNPs were identified by alignments to the scaffolds of the *A. duranensis* and *A. ipaensis* genome assemblies as well as local realignment and probabilistic variant calling in CLC Genomics Workbench (CLC Bio). Filtering in CLC Workbench resulted in about 3 million high-quality homozygous-parental SNPs for both A- and B-genome mapping population parents. The coordinates of these SNPs were converted into BED format, and the alignment data at the SNP coordinates were extracted with SAMtools mpileup60. From the low-coverage sequencing data, groups of 20 consecutive SNPs were haplotyped with a set of custom Python scripts. Genotype calls were inspected visually and by a hidden Markov model (HMM) script (courtesy of Ian Korf, University of California, Davis) to identify population-wide switches in genotype calls corresponding to scaffold misjoins. Scaffolds not displaying recombination for an individual RIL were haplotyped. Linkage groups were identified from the haplotyping data using MadMapper and Carthagene, applying logarithm of odds (LOD) score thresholds of 8 and distance thresholds of 50 cM; genetic maps were generated with Carthagene using the lkh traveling salesman algorithm and flips, polish and annealing optimizations. Additional scaffolds (indicated in the data files) were added to genetic bins in two rounds of binning with a custom Python script. Misjoined scaffolds were split at breakpoint locations identified by flanking GBS SNP locations, at the "upstream SNP" and the "downstream SNP", delineating the switches in genotype calls, and intervening sequence was excluded from the pseudomolecule assembly.

### 9.4.3 Generation of Chromosomal Pseudomolecules

Under the IPGI project, scaffolds less than 10 kb in length were removed (they are available in the full assembly scaffold files at PeanutBase: Adur1.split6.fa and Aipa2 s.split7.fa, http://peanutbase.org/files/genomes/Arachis_ipaensis/assembly/). Sequences were subjected to RepeatMasker using *Arachis* repeat libraries available at PeanutBase (mobile-elements-AA051914.fasta and mobile-elements-BB051914.fasta). Pseudomolecules were given initial chromosomal placements and orderings according to the GBS maps. Placement was arbitrary within blocks with the same centiMorgan value. Scaffold orientation and placement were refined according to the different genetic maps such as the tetraploid AB-genome map, the diploid

A-genome map (for the *A. duranensis* assembly), the diploid B-genome map (for the *A. ipaensis* assembly) and finally the tetraploid AB-genome consensus map (Shirasawa et al. 2013). Markers were located on the scaffolds using BLAST and ePCR (electronic PCR) with high similarity parameters (taking the top hits only, with placement by BLAST (*e* value < 1 × 10 − 10) given preference over ePCR where both were available). Markers placing scaffolds on linkage groups other than the one assigned by the GBS data were dropped.

Where allowed by map data, scaffold positions and orientations were adjusted using synteny between the two *Arachis* species and, where necessary (generally within pericentromeric regions), synteny with *G. max* and *Proteus vulgaris*; the presence of telomeric repeats near chromosome ends; information from repeat-masked paired-end sequences from 42,000 BAC clones of *A. duranensis* V14167 (FI321525–FI281689) and Moleculo sequence reads from *A. ipaensis* and *A. duranensis*. Apparent inversions were visually inspected and confirmed. Scaffolds with either <5000 non-N bases or <20,000 bp in length and with <10,000 non-N bases were removed. Pseudomolecules were generated with 10,000 Ns separating the scaffold sequences and were oriented and numbered in accordance with previously published maps (Shirasawa et al. 2013; Gautami et al. 2012; Moretzsohn et al. 2005 and 2009). The scaffolds were thus assigned to pseudomolecules under IPGI project. Due to unavailability of proper information on linkage mapping on A-genome, the genome assembly was made at scaffold level under DPPAGSC project.

### 9.4.4 Gene Prediction and Annotation

Under IPGI project, genome assemblies were masked with RepeatMasker using the repeat libraries developed for the two diploid species and annotated for gene models using the MAKER-P pipeline (Campbell et al. 2014). *Arachis*-specific models for the ab initio gene predictor SNAP were trained using high-scoring gene models from a first iteration of the pipeline and then used in the final annotation pass; no training was done for the other ab initio predictors included in the pipeline. RNA sequencing de novo assemblies for *A. hypogaea* and the diploid *Arachis* species were supplied as transcript evidence along with available EST and mRNA data sets from NCBI for these same species. Further evidence was supplied by proteomes derived from the annotations for *G. max*, *P. vulgaris,* and *Medicago truncatula* as represented in Phytozome v. 10. Default MAKER-P parameters were used for all other options, with the exception of disabling splice isoform prediction (alt_splice = 0) and forcing start and stop codons into every gene (always_complete = 1). The resulting MAKER-P gene models were post-processed to exclude from the main annotation files gene models with relatively poor support (annotation evidence distance scores of ≥0.75) or with significant BLASTN homology to identified mobile-elements (HSP (high-scoring segment pair) coverage over ≥50% of the transcript sequence at ≥80% identity and *e* value ≤1 × 10 − 10). Provisional functional assignments for the gene models were produced using InterProScan and BLASTP against annotated proteins from *Arabidopsis thaliana*, *G. max,* and *M. truncatula*, with outputs processed using AHRD (https://github.com/groupschoof/AHRD), for lexical analysis and selection of the best functional descriptor of each gene product.

Under DPPAGSC project, to annotate the *A. duranensis* genome, an automated genome annotation pipeline MAKER was used that aligns and filters EST and protein homology evidence and produces de novo gene prediction, infers 5′ and 3′ UTR, and integrates these data to generate final downstream gene models with quality control statistics. Several iterative runs of MAKER were used to produce the final gene set. In total, 50,324 gene models for *A. duranensis* were predicted. All predicted protein sequences were functionally annotated using the BLAST + (version 2.2.27) with a threshold E-value of 1e-5 against a variety of protein and nucleotide databases, including the NCBI

nucleotide (NT), the non-redundant protein (NR), the Conserved Domain Database (CDD), the UniProtKB (www.uniprot.org), Pfam and the Gene Ontology (GO). The *A. duranensis* genes were also mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway maps of KEGG databases. To infer functions for the predicted genes, InterProScan was used to search the predicted genes against the protein signature from InterPro with default parameters. Fifteen gene sets from legumes, oilseed crops and other plant species were used for comparative analysis. A Cytoscape plugin BiNGO was used for enrichment analysis with hypergeometric test and Benjamini multiple testing correction at a significance level of 0.01.

## 9.5 Assembly of Diploid Genomes

The total assembly sizes were 1.21 and 1.51 Gb for *A. duranensis* and *A. ipaensis*, respectively, from the data generated from the seven paired-end libraries corresponded to an estimated 154X and 163X base-pair coverage from IPGI (Table 9.1). The assembly size of *A. duranensis* obtained by DPPAGSC is 1.05 Gb with 57.14X read depth (Table 9.1).

The approximate chromosomal/pseudomolecule placements were obtained by using ultradense genetic maps in case of IPGI project. SNPs within scaffolds were used to validate the assemblies and confirmed their high quality; 190 of 1297 initial scaffolds of *A. duranensis* and 49 of 353 initial scaffolds of *A. ipaensis* were identified as chimeric, on the basis of the presence of diagnostic population-wide switches in genotype calls occurring at the point of misjoin. Overall, 96.0 and 99.2% of the sequence in contigs $\geq$ 10,000 bp in length, represented by 1692 and 459 scaffolds, could be ordered into 10 chromosomal pseudomolecules per genome of 1025 and 1338 Mb for *A. duranensis* and *A. ipaensis*, respectively (Aradu.A01–Aradu.A10 and Araip.B01–Araip.B10; GenBank, assembly accessions GCA_000817695.1 and GCA_000816755.1. The pseudomolecules mostly showed one-to-one equivalence between the A- and B-genomes and were numbered according to previously published linkage maps (Shirasawa et al. 2013; Gautami et al. 2012; Moretzsohn et al. 2005, 2009). Comparisons of the chromosomal pseudomolecules with 14 BAC sequences from *A. duranensis* and 6 BAC sequences from *A. ipaensis* showed collinearity of contigs and high sequence identity ($\geq$ 99%).

Whereas in China-led initiative, PCR amplification of randomly selected regions, sequence-depth distribution, and expressed sequence tag validation indicated the high quality of the assembled genome with 8173 scaffolds. K-mer analysis indicated *A. duranensis* genome size of 1.38 Gb that is consistent with previous report (Temsch and Greilhuber 2000). However 50,324 protein coding gene models were predicted using transcriptome sequences (Table 9.1). When compared with the gene sets of legumes, oilseeds, and other plant species, *A duranensis* showed highest similarity to legumes with gene numbers comparable with *Medicago truncatula* (50,894), lower than soybean (tetraploid Glycine max, 56,044), and higher than other legumes.

### 9.5.1 Repetitive Sequences

Under the IPGI project, the transposable elements accounted for 61.7 and 68.5% of the *A. duranensis* and *A. ipaensis* genomes respectively with long terminal repeat (LTR) comprise of more than 50% of each genome. This observation was similar in DPPAGSC study as well where about 59.77% of the *A. duranensis* genome appeared to have transposable elements with ~40% LTR retrotransposons. These observations were comparable with the estimated repetitive content (64%) for cultivated peanut using renaturation kinetics in the past (Dhillon and Rake 1980). The DNA transposons constituted about 10% of the genome under IPGI project whereas they were about 5.19% in case of *A. duranensis* under DPPAGSC project. The long interspersed nuclear elements (LINEs) were about 7.8 and 11.7% in *A. duranensis* and *A. ipaensis* genomes respectively (US-led initiative) and only about 1.26% of the *A. duranensis* genome (China-led initiative). Besides, under

DPPAGSC project, a total of 105,003 simple sequence repeats (SSRs) were identified in *A. duranensis*. Furthermore, resequencing of two other A-genome genotypes and four B-genome genotypes allowed the discovery of ~8 million SNPs and other structural variations.

### 9.5.2 Gene Annotation and Analysis of Gene Duplications

Under IPGI project, transcript assemblies were constructed using sequences expressed in diverse tissues of *A. duranensis* V14167, *A. ipaensis* K30076, and *A. hypogaea* cv. Tifrunner (16,439,433, 21,406,315, and 2,064,268,316 paired-end reads for each species, respectively. Using these assemblies and representative characterized transposon sequences, 36,734 and 41,840 high-quality non-transposable element genes for *A. duranensis* and *A. ipaensis*, respectively were generated (Table 9.1). The elevated gene numbers in *A. ipaensis* appear to originate from more local duplications, which can be seen in counts of genomically "close" paralogous genes. Considering similar genes within a ten-gene window, there were 25% more in *A. ipaensis* than in *A. duranensis* (7825 vs. 6241). Gene families known to occur in clusters such as those encoding NB-ARC, leucine-rich repeat (LRR), pentatricopeptide-repeat, kinase, WD40-repeat, and kinesin proteins had large differential counts between the two genomes. These differences were also apparent with wider inspection. In a set of 9236 gene families with members in *A. ipaensis* or *A. duranensis*, or both, 2879 families had more members in *A. ipaensis*, 1983 had more members in *A. duranensis* and 4374 had the same number of members in both species.

Under DPPAGSC project, about 50,324 protein coding gene models were predicted using transcriptome sequences in *A. duranensis*. When compared with the gene sets of legumes, oilseeds, and other plant species, *A duranensis* showed highest similarity to legumes with gene numbers comparable with *Medicago truncatula* (50,894), lower than soybean (tetraploid *Glycine max*, 56,044), and higher than other legumes. Of the 50,324 gene models, ~90% matched entries in publically available databases. Approximately 10.9% (5494) of gene models with no homology to known proteins were supported by transcriptome data and may be peanut-specific. A total of 5251 putative *A. duranensis* transcription factor genes in 57 families, 10.4% of the predicted *A. duranensis* genes, slightly higher than soybean, and much higher than most plant species were analyzed. Certain TFs like B3, E2F/DP, FAR1, GeBP, HSF, NAC, S1Fa-like, and STAT were dominant in *A. duranensis.* Families such as ARR-B, CAMTA, DBB, MIKC, and NF-YA, were sparser in *A. duranensis* than in most plants. Expansion and contraction of TF families may reflect regulatory differences in biological functions of *A. duranensis*. In this study, 816 *Arachis* microRNAs (miRNAs), 913 transfer RNAs (tRNAs), 115 ribosomal RNAs (rRNAs), and 202 small nucleolar RNAs (snRNAs) were also annotated. A total of 64 target genes were predicted after aligning 15 new miRNAs to gene models.

### 9.5.3 Gene Evolution and Genome Duplication

The IPGI project analyses suggest that the *Arachis* lineages have been accumulating mutations relatively quickly since the divergence of the Dalbergioid clade ~58 million years ago. Modal *K*S values (synonymous substitutions per synonymous site) for paralogs are approximately 0.95 for *A. ipaensis* and 0.90 for *A. duranensis*, more similar to that the Ks value for *Medicago* paralogs of ~0.95) than to those of *Lotus* (~0.65), *Glycine* (~0.65) or *Phaseolus* (~0.80). Average rates of change for *Arachis* genes were estimated at $8.12 \times 10 - 9$ *KS*/year. *Arachis* has accumulated silent changes at a rate ~1.4 times faster than that in *G. max*. On the basis of average rates of change for *Arachis* of $8.12 \times 10^{-9}$ *KS*/year, it was estimated that *A. duranensis* and *A. ipaensis* diverged ~2.16 million years ago.

Under the DPPAGSC project, the genome duplication of *A. duranensis* was compared with

that of *Medicago* and soybean. Collinear genes from *Medicago*, soybean (*Glycine max*), and grape (*Vitis vinifera*) were used to analyze related evolutionary events. The Ks distribution of peanut homologs shows a prominent peak around Ks = 0.5, overlapping the peak of soybean duplicated genes resulting from a pan-legume tetraploidization previously inferred to be ~60 Mya (Young et al. 2011). Adding the pan-eudicot γ-hexaploidy (~130 Mya) and polyploidy producing tetraploid peanut by joining the *Arachis* A and B subgenomes, estimated to have diverged 3.5 Mya (Nielen et al. 2012), the *Arachis* lineage has been affected by at least three polyploidizations since the origin of eudicots, with a collective 12X paleoduplication depth.

In addition, the gene conversion among the subgenomes was discussed in the DPPAGSC study, where there is unidirectional homeologous exchanges between genes from different subgenomes can overwrite one progenitor allele with additional copies of the other (Paterson et al. 2012; Wang et al. 2012). Implicated as a possible contributor to the transgressive properties of polyploids relative to their progenitors, extensive gene conversion was inferred to have occurred about 7500–12,500 years ago since formation of the Neolithic species *Brassica napus* (Chalhoub et al. 2014). By performing a three-way comparison of the synthetic tetraploid ISATGR 184 and its progenitor lines, ICG 8123 and ICG 8206, evidence of extensive gene conversion was observed between subgenomes in the ~ three seed-to-seed generations since its formation by human hands. The vast majority (~93%) of alleles have been converted to homozygosity for the A-genome allele in ISATGR 184, an asymmetry resembling those found in cotton and canola (Young et al. 2011, Chalhoub et al. 2014). ISATGR 1212, a reciprocal cross between the same parental lines as ISATGR 184, shared Bt to At bias of conversion but had far fewer converted sites than ISATGR 184 ($\chi 2 << 0.001$), perhaps indicating a contribution of germ-line types to genomic variation in the offspring.

## 9.6 Synteny with Allied and Model Genomes

IPGI study provided the syntenic relations between A and B subgenomes and their sequence comparison with tetraploid peanut. Most pseudomolecules had symmetrically positioned pericentromeres that was in accordance with cytogenetic observations (Robledo and Seijo 2010; Robledo et al. 2009). Most pseudomolecules showed a one-to-one correspondence between the two species: pairs 02, 03, 04, and 10 were collinear; pairs 05, 06, and 09 were each differentiated by a large inversion in one arm of one of the pseudomolecules; and the pseudomolecules in pair 01 were differentiated by large inversions of both arms. In contrast, chromosomes 07 and 08 have undergone complex rearrangements that transported repeat-rich DNA to A07 and gene-rich DNA to A08. As a result, A07 has only one normal (upper) euchromatic arm and A08 is abnormally small, with low repetitive content. In accordance with cytogenetic observations (Seijo et al. 2007; Nielen et al. 2010), A08 could be assigned as the characteristic small "A chromosome" (cytogenetic chromosome A09).

All *A. ipaensis* pseudomolecules were larger than their *A. duranensis* counterparts. This is partly because of a greater frequency of local duplications and higher transposon content in *A. ipaensis*. In chromosomes without inversions, there were characteristic density gradients for genes, repetitive DNA and methylation (with gene densities increasing and densities of repetitive DNA and methylation decreasing toward chromosome ends). However, in regions that had undergone large rearrangements, in *A. duranensis*, these gradients were disrupted. From these observations, we concluded that most major rearrangements occurred in the A-genome lineage. Size differences between homeologous chromosomes that were differentiated by large rearrangements tended to be greater than those between collinear ones. Because the *A. duranensis* chromosomes that have undergone inversions are smaller than expected, it is evident that,

in this dynamic, on balance, the elimination of DNA has predominated over its accumulation. Comparisons with *Phaseolus vulgaris* L., which shared a common ancestor with *Arachis* about 58 million years ago, showed syntenous chromosomal segments. In some cases, there was almost a one-to-one correspondence between chromosomes (for example, B01 and Pv03, B05 and Pv02, B06 and Pv01, and B08 and Pv05).

Sequence comparison to tetraploid peanut showed fundamentally one-to-one correspondences between the diploid chromosomal pseudomolecules and cultivated peanut linkage groups. Of the marker sequences from three maps (Shirasawa et al. 2012; Zhou et al. 2014), 83, 83, and 94% were assigned by sequence similarity searches to the expected diploid chromosomal pseudomolecules. For more detailed genome-wide comparisons, about 5.74 Gb (2X coverage) of long-sequence Moleculo reads from *A. hypogaea* cv. Tifrunner were generated and mapped the reads to the combined diploid pseudomolecules. The corrected median identities between the *A. hypogaea* Moleculo reads and the pseudomolecules of *A. duranensis* and *A. ipaensis* were 98.36 and 99.96%, respectively. When visualized as plots along the chromosomal pseudomolecules, the diploid A-genome chromosomes were distinctly less similar to *A. hypogaea* sequences than the B-genome chromosomes.

## 9.7 Trait Understanding

DPPAGSC project also provided insights into some unique traits found in peanut-like fructification, oil biosynthesis, and allergens. A unique characteristic of peanut is the peg/gynophore, a specialized organ that grows downwards upon fertilization, driving the developing pod into the soil. Fruit development in other plants is controlled in light; on the contrary there is subterranean fructification in peanut. A total of 151 genes related to "gravitropism" were found during pod development. Five TF families related to photomorphogenesis were identified in very large numbers in *A. duranensis*, namely S1Fa-like, FAR1, HSF, NAC, and STAT. S1Fa-like TFs containing a small peptide (70 aa) with a nuclear localization and DNA binding domain were more highly expressed in roots and etiolated seedlings than green leaves. The FAR1 TF family plays an important role in modulating phyA-signaling homeostasis in higher plants (Lin et al. 2007). Importantly, phyA localized in the cytosol of dark-grown seedlings acts primarily as a far-red sensor, which regulates the transition from skotomorphogenesis to photomorphogenesis (Whitelam and Halliday 2008). PhyB, exhibiting a fast and strong but incomplete dark conversion in some cases, is the main light receptor responsible for the shade-avoidance response in mature plants (Medzihradszky et al. 2013) and shows evidence of positive selection in *A. duranensis* suggesting a role in skotomorphogenesis.

Oil biosynthesis is one more important trait of peanut, where better understanding of this trait will be very helpful to breed confectionary suitable peanut. Considering the importance of peanut as an oil crop, annotations of 67 gene models were searched for their similarity with the genes involved in fatty acid biosynthesis and triacylglycerol (TAG), that represent the oleic and linoleic acids (Moore and Knauft 1989). FAD2 encoding δ-12 oleic acid desaturase, the key enzyme controlling the high oleate trait, was highly expressed in seed filling but less during desiccation. Genes encoding key enzymes in the TAG pathway were expressed at diverse levels at different developmental stages. Multiple copies or isoforms of some key genes were detected in the *A. duranensis* genome like glycerol-3-phosphate acyltransferase and diacylglycerol acyltransferase, which catalyze the first and final steps in the TAG pathway. Information on copy number and expression diversity of these metabolic genes is important for improvement of oil quality parameters in peanut, such as a high oleic to linoleic acid ratio (O/L).

Peanut allergy is one of the most serious life-threatening food sensitivities prevalent among a section of world population, particularly

among children. Comparison with known allergenic proteins from peanut and other crops identified 21 candidate allergen-encoding genes in *A. duranensis*, of which nine have already been reported in peanut and others are homologs from other crops. Understanding of allergen-encoding genes in peanut can be utilized to produce allergy-free peanuts either by genomics-enabled breeding or by cis-genic approaches.

## 9.8 Genome Dominance

Whole-genome duplications have occurred in many eukaryotic lineages, particularly in plants. Following most ancient tetraploidies, the two subgenomes are distinguishable, because the dominant subgenome tends to have more genes than the other subgenome. Additionally, among retained pairs, the gene on the dominant subgenome tends to be expressed more than its recessive homeolog (Woodhouse et al. 2014).

The most thorough study of the location and number of rDNAs was conducted by Seijo and collaborators (2004) using fluorescent in situ hybridization (FISH). The study showed, as previously mentioned, that the number, size, and distribution of rDNA clusters in *A. hypogaea* are virtually equivalent to the sum of those present in *A. duranensis* and *A. ipaënsis*. A single pair of 5S sites is present on each of the A and B chromosome complements, and two pairs of 18S-25S sites on the A chromosomes and three pairs on the B. The only exception to this equivalence is that in both of the diploid species, 18S-25S sites bear a thread-like constriction indicating intense transcriptional activity (forming the SAT chromosome; Fernandez and Krapovickas 1994). However, in the allotetraploid the constrictions are observed only on the A-genome. This indicates that the transcriptional activity of the B-genome rDNAs has been silenced, a common event in polyploids called nucleolar dominance (Cermeno et al. 1984; Preuss and Pikaard 2007).

## 9.9 Conclusion

The IPGI project has used the genome information to identify candidate pest- and disease-resistance genes, to reduce collapse in tetraploid transcriptome assemblies and to show the impact of recombination between subgenomes in cultivated peanut. Besides providing basic knowledge about the A-genome (*A. duranensis)* progenitor, the DPPAGSC project also provided a major source of candidate genes for fructification, oil biosynthesis, and allergens, expanding knowledge of understudied areas of plant biology and human health impacts of plants. The availability of these genomes will lead to further advances in knowledge of genetic changes since the very recent polyploidization event that gave rise to cultivated peanut and to the production of better tools for molecular breeding and crop improvement.

## References

Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EK, Liu X, Gao D, Clevenger J, Dash S, Ren L et al (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet 48(4):438–446

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W (2011) Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27(4):578–579

Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, Lei J, Achawanantakun R, Jiao D, Lawrence CJ, Ware D et al (2014) MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. Plant Physiol 164(2):513–524

Cermeno MC, Orellana J, Santos JL, Lacadena JR (1984) Nucleolar activity and competition (amphiplasty) in the genus *Aegilops*. Heredity 53(3):603–611

Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Corréa M et al (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science 345(6199):950–953

Chen X, Li H, Pandey MK, Yang Q, Wang X, Garg V, Li H, Chi X, Doddamani D, Hong Y, Upadhyaya H et al (2016) Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into

geocarpy, oil biosynthesis, and allergens. Proc Natl Acad Sci 113(24):6785–6790
Dhillon SS, Rake AV (1980) Miksche JP. Reassociation kinetics and cytophotometric characterization of peanut (*Arachis hypogaea* L.) DNA. Plant Physiol 65 (6):1121–1127
Fávero AP, Simpson CE, Valls JF, Vello NA (2006) Study of the evolution of cultivated peanut through crossability studies. Crop Sci 46(4):1546–1552
Fernández A, Krapovickas A (1994) Cromosomas Y Evolucion En" *Arachis* (Leguminosae)". *Bonplandia* 187–220
Foncéka D, Hodo-Abalo T, Rivallan R, Faye I, Sall MN, Ndoye O, Fávero AP, Bertioli DJ, Glaszmann JC, Courtois B, Rami JF (2009) Genetic mapping of wild introgressions into cultivated peanut: a way toward enlarging the genetic basis of a recent allotetraploid. BMC Plant Biol 9(1):103
Gautami B, Foncéka D, Pandey MK, Moretzsohn MC, Sujay V, Qin H, Hong Y, Faye I, Chen X, BhanuPrakash A, Shah TM et al (2012) An international reference consensus genetic map with 897 marker loci based on 11 mapping populations for tetraploid groundnut (*Arachis hypogaea* L.). PLoS One 7(7):e41213
Huang S, Chen Z, Huang G, Yu T, Yang P, Li J, Fu Y, Yuan S, Chen S, Xu A (2012) HaploMerger: reconstructing allelic relationships for polymorphic diploid genome assemblies. Genome Res 22(8):1581–1588
Jackson SA (2016) Rice: The first crop genome. Rice. doi:10.1186/s12284-016-0087-4
Jia J, Zhao S, Kong X, Li Y, Zhao G, He W, Appels R, Pfeifer M, Tao Y, Zhang X, Jing R et al (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496 (7443):91–95
Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Li S et al (2010) De novo assembly of human genomes with massively parallel short read sequencing. Genome Res 20(2):265–272
Lin R, Ding L, Casola C, Ripoll DR, Feschotte C, Wang H (2007) Transposase-derived transcription factors regulate light signaling in *Arabidopsis*. Science 318(5854):1302–1305
Ling HQ, Zhao S, Liu D, Wang J, Sun H, Zhang C, Fan H, Li D, Dong L, Tao Y, Gao C et al (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. Nature 496(7443):87–90
Liu B, Yuan J, Yiu SM, Li Z, Xie Y, Chen Y, Shi Y, Zhang H, Li Y, Lam TW, Luo R (2012) COPE: an accurate k-mer-based pair-end reads connection tool to facilitate genome assembly. Bioinformatics 28 (22):2870–2874
Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience. 1(1):1
Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M International wheat genome sequencing consortium, et al. (2014) Ancient hybridizations among the ancestral genomes of bread wheat. Science 345
McCoy RC, Taylor RW, Blauwkamp TA, Kelley JL, Kertesz M, Pushkarev D, Petrov DA, Fiston-Lavier AS (2014) Illumina TruSeq synthetic long-reads empower de novo assembly and resolve complex, highly-repetitive transposable elements. PLoS One 9 (9):e106689
Medzihradszky M, Bindics J, Adam E, Viczian A, Klement E, Lorrain S, Gyula P, Merai Z, Fankhauser C, Medzihradszky KF, Kunkel T, Schafer E, Nagy F (2013) Phosphorylation of phytochrome B inhibits light-induced signaling via accelerated dark reversion in Arabidopsis. The Plant Cell 25(2):535–544
Michael TP, Jackson S (2013) The first 50 plant genomes. The Plant Genome, 6(2)
Moretzsohn MC, Leoi L, Proite K, Guimaraes PM, Leal-Bertioli SC, Gimenes MA, Martins WS, Valls JF, Grattapaglia D, Bertioli DJ (2005) A microsatellite-based, gene-rich linkage map for the AA genome of Arachis (Fabaceae). Theor Appl Genet 111(6):1060–1071
Moretzsohn MC, Barbosa AV, Alves-Freitas DM, Teixeira C, Leal-Bertioli SC, Guimarães PM, Pereira RW, Lopes CR, Cavallari MM, Valls JF, Bertioli DJ (2009) A linkage map for the B-genome of *Arachis* (Fabaceae) and its synteny to the A-genome. BMC Plant Biol 9(1):40
Moore KM, Knauft DA (1989) The inheritance of high oleic acid in peanut. J Heredity 80:252–253
Moretzsohn MC, Gouvea EG, Inglis PW, Leal-Bertioli SC, Valls JF, Bertioli DJ (2013) A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. Ann Bot 111 (1):113–126
Nielen S, Campos-Fonseca F, Leal-Bertioli S, Guimarães P, Seijo G, Town C, Arrial R, Bertioli D (2010) FIDEL—a retrovirus-like retrotransposon and its distinct evolutionary histories in the A-and B-genome components of cultivated peanut. Chromosome Res 18 (2):227–246
Nielen S, Vidigal BS, Leal-Bertioli SC, Ratnaparkhe M, Paterson AH, Garsmeur O, D'Hont A, Guimaraes PM, Bertioli DJ (2012) Matita, a new retroelement from peanut: characterization and evolutionary context in the light of the Arachis A-B genome divergence. Mol Genet Genomics 287(1):21–38
Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J, Yoo MJ (2012) Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature 492(7429):423–427
Preuss S, Pikaard CS (2007) rRNA gene silencing and nucleolar dominance: insights into a chromosome-scale epigenetic on/off switch. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression 1769(5):383–92

Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, Cheng J, Zhao S, Xu M, Luo Y, Yang Y et al (2014) Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. Proc Natl Acad Sci 111 (14):5135–5140

Ramos ML, Fleming G, Chu Y, Akiyama Y, Gallo M, Ozias-Akins P (2006) Chromosomal and phylogenetic context for conglutin genes in *Arachis* based on genomic sequence. Mol Genet Genomics 275(6):578–592

Robledo G, Seijo G (2010) Species relationships among the wild B genome of Arachis species (section Arachis) based on FISH mapping of rDNA loci and heterochromatin detection: a new proposal for genome arrangement. Theor Appl Genet 121(6):1033–1046

Robledo G, Lavia GI, Seijo G (2009) Species relations among wild *Arachis* species with the A genome as revealed by FISH mapping of rDNA loci and heterochromatin detection. Theor Appl Genet 118 (7):1295–1307

Samoluk SS, Chalup L, Robledo G, Seijo JG (2015) Genome sizes in diploid and allopolyploid Arachis L. species (section *Arachis*). Genet Resour Crop Evol 62 (5):747–763

Schatz MC, Witkowski J, McCombie WR (2012) Current challenges in de novo plant genome sequencing and assembly. Genome Biol 13(4):1

Seijo JG, Lavia GI, Fernandez A, Kaprovickas A, Ducasse D, Moscone EA (2004) Physical mapping of the 5S and 18S-25S RRNA gene by FISH as evidence that Arachis duranensis and A. ipaensis are the wild diploid progenitors of *A. hypogaea* (Leguminosae). Am J Bot 91:1294–1303

Seijo G, Lavia GI, Fernández A, Krapovickas A, Ducasse DA, Bertioli DJ, Moscone EA (2007) Genomic relationships between the cultivated peanut (*Arachis hypogaea*, Leguminosae) and its close relatives revealed by double GISH. Am J Bot 94 (12):1963–1971

Shirasawa K, Koilkonda P, Aoki K, Hirakawa H, Tabata S, Watanabe M, Hasegawa M, Kiyoshima H, Suzuki S, Kuwata C, Naito Y (2012) In silico polymorphism analysis for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. BMC Plant Biol 12(1):1

Shirasawa KE, Bertioli DJ, Varshney RK, Moretzsohn MC, Leal-Bertioli SC, Thudi MA, Pandey MK, Rami JF, Foncéka DA, Gowda MV, Qin HO et al (2013) Integrated consensus map of cultivated peanut and wild relatives reveals structures of the A and B genomes of Arachis and divergence of the legume genomes. DNA Res 20(2):173–184

Soltis DE, Visger CJ, Soltis PS (2014) The polyploidy revolution then and now: Stebbins revisited. Am J Bot 101(7):1057–1078

Temsch EM, Greilhuber J (2000) Genome size variation in *Arachis hypogaea* and *A. monticola* re-evaluated. Genome 43(3):449–451

Varshney RK, Nayak SN, Jackson S, May G (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27(9):522–530

Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S, Zou C et al (2012) The draft genome of a diploid cotton *Gossypium raimondii*. Nat Genet 44(10):1098–1103

Whitelam GC, Halliday KJ (2008) Annual plant reviews, light and plant development. John Wiley & Sons

Woodhouse MR, Cheng F, Pires JC, Lisch D, Freeling M, Wang X (2014) Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. Proc Natl Acad Sci 111(14):5283–5288

Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KF, Gouzy J, Schoof H, Van de Peer Y et al (2011) The Medicago genome provides insight into the evolution of rhizobial symbioses. Nature 480(7378):520–524

Zhou X, Xia Y, Ren X, Chen Y, Huang L, Huang S, Liao B, Lei Y, Yan L, Jiang H (2014) Construction of a SNP-based genetic linkage map in cultivated peanut based on large scale marker development using next-generation double-digest restriction-site-associated DNA sequencing (ddRADseq). BMC Genom 15(1):1

# Functional Genomics in Peanut Wild Relatives

10

Patricia M. Guimarães, A.C.M. Brasileiro, A. Mehta and A.C.G. Araujo

**Abstract**

Breeding efforts to develop peanut (*Arachis hypogaea* L.) cultivars with multiple resistances, high seed quality and yield have been hindered by the narrow genetic base of the cultivated genepool and the complex nature of its genome. Conversely, peanut wild relatives (*Arachis* spp.) are considered important sources of resistance alleles, since they have high genetic diversity and have been selected during evolution in a range of adverse environments and conditions. Transcriptome studies on wild species constitute important assets for the identification of genomic segments of interest for transfer into cultivated species. An unprecedented amount of genomic information for wild and cultivated *Arachis* has been produced in recent years, leading to the discovery of genes and regulatory sequences, and enlarging the collections of molecular markers. The increasing availability of *Arachis* transcriptomic resources such as ESTs, Unigenes, full-length cDNA clones and derived proteins is enabling a more precise correlation of genotype/phenotype in the genus, with the potential to facilitate accurate intervention in pathways to improve peanut agronomical traits. To maximize these valuable assets, candidate gene validation and peanut genetic transformation methods have been developed to facilitate the deployment of wild alleles into new cultivars.

P.M. Guimarães (✉) · A.C.M. Brasileiro · A. Mehta · A.C.G. Araujo
Embrapa Genetic Resources and Biotechnology (Cenargen) - Parque Estação Biológica (PqEB), Av. W5 Norte (Final), P.O. Box 0237270-70770-917, Brasília, DF, Brazil
e-mail: patricia.guimaraes@embrapa.br

A.C.M. Brasileiro
e-mail: ana.brasileiro@embrapa.br

A. Mehta
e-mail: angela.mehta@embrapa.br

A.C.G. Araujo
e-mail: ana-claudia.guerra@embrapa.br

## 10.1 Introduction

Time and place of legumes origin are not yet clear, although there are molecular phylogenetic reports suggesting that their dispersion date from the tertiary age (Wojciechowski 2003). Leguminosae is the third largest family in angiosperms, with several species constituting important sources of proteins for human diet and livestock forage, oil, fiber and green manure. Additionally,



R.K. Varshney et al. (eds.), *The Peanut Genome*, Compendium of Plant Genomes, DOI 10.1007/978-3-319-63935-2_10

due to the ability of many species to form nitrogen-fixing symbioses with rhizobia (Sprent et al. 2009), they also play a strategic role in sustainable agriculture. Cultivated legumes, similarly to other crops, generally show a reduced genetic diversity when compared to their wild relatives. This effect is basically related to small population sizes "founder effect" that occur during the initial formation of a domesticated lineage, added to directional selection for alleles associated with evolution and domestication (Abbo et al. 2012; Gepts 2014).

*Arachis* belongs to the legume subfamily Papilionoid, within the Dalbergioid clade, and has its center of diversity at the eastern Andean area, comprising Bolivia, Argentina, Paraguay and Brazil (Krapovickas et al. 2007; Kochert et al. 1996; Valls and Simpson 2005). Within the genus, the largest section holds the only cultivated species, *A. hypogaea* L., known as peanut or groundnut. In this section, except for peanut and the wild tetraploid *A. monticola*, all species have diploid genomes (Krapovickas 1994), a ploidy barrier which has contributed to the isolation of the cultivated species.

Peanut is a recent allopolyploid (~3500 years old) (Bonavia 1982), containing two different genomic components, the A and B, highly conserved, complex and organized in 40 chromosomes (Husted 1936; Singh and Moss 1984; Kochert et al. 1996). The origin of these A and B components is most probably, a unique successful hybridization between two wild diploids, *A. duranensis* (A genome) and *A. ipaënsis* (B genome), followed by a natural duplication of the chromosomes that restored the fertility of the hybrid (Seijo et al. 2004). Thus, a severe genetic bottleneck was imposed at the origin, which was enlarged by cross- incompatibility with other wilds, preferential self-pollination, cultivation processes and variation in the dynamics of populations. This created the conditions to develop a genetic drift, genetic isolation and species divergence which cumbered peanut breeding (Bertioli et al. 2014).

Targeted traits for peanut improvement are focused on food quality and flavor, higher yields and the production of seeds free from mycotoxins (Janila et al. 2013). Nevertheless, most cultivars are susceptible to various biotic and abiotic stresses, such as fungal foliar diseases, nematodes, and drought that can limit its productivity and seed quality (Devi et al. 2011). On the other hand, a number of wild *Arachis* species encompass desirable alleles for several economically important traits, such as resistance to diseases and insect pests (Dwivedi et al. 2003; Rao 2003; Kalyani et al. 2007; Nautiyal et al. 2008; Pandey et al. 2012; Singsit et al. 1995) and tolerance to dry conditions (Brasileiro et al. 2015; Leal-Bertioli et al. 2012). Therefore, genome-wide introgression of genomic segments of wild species into selected peanut background is an interesting approach to explore valuable alleles available in the wilds (Stalker et al. 2013), even though possible effects of genetic linkage drag might confer unattractive traits to cultivars (Sharma et al. 2013). It is also important to note that, to enable the full use of wild alleles for peanut breeding, further advances are essential on germplasm characterization, breeding tools, large-scale phenotyping and escalation of "omics" data analysis (Mochida et al. 2015).

The first peanut improved variety using wild alleles was COAN (Simpson and Starr 2001), obtained by the introgression of a segment of a single chromosome from *A. cardenasii* (Nagy et al. 2010), that conferred resistance to the root-knot nematode (RKN) (*Meloidogyne* spp.) (Holbrook and Stalker 2003). Following breeding with COAN and other peanut cultivars, new varieties (NemaTAM, Tifguard, Tirupati, GPBD4, Webb, High O/L etc.) were produced, showing also improved resistance to fungal diseases, oil content, and other traits. However, *A. cardenasii* is the only RKN resistance source currently in use, thus making the identification of new resistance sources critical to assure breeding advances.

More recently, the characterization of a number of wild species harboring resistances/tolerance to biotic and abiotic stresses have opened new opportunities for peanut improvement, as they show resistance to RKN (Proite et al. 2007; Leal-Bertioli et al. 2009; Guimaraes et al. 2015; Proite et al. 2008), web

blotch and scab (Reddy et al. 2000; Sharma et al. 1999; Pande and Rao 2001; Guimaraes et al. 2012; Michelotto et al. 2015; Leal-Bertioli et al. 2009), and also show a more suitable performance under drought stress (Brasileiro et al. 2015; Leal-Bertioli et al. 2012). These wild species, therefore, have great potential to contribute as sources of resistance/tolerance alleles for peanut improvement.

To identify these novel alleles or the expression variants present in peanut wild relatives, the application of massal transcriptome sequencing has been a powerful tool. This has helped to characterize genes related to agronomic traits, allelic expression differences and further understanding of genes role, genome organization and evolution, genetic variations and genomic basis of wilds adaptation and peanut domestication and diversification (Brozynska et al. 2015).

Although grain legumes are the second most important human dietary source, information on the genome, transcriptome, proteome and other "omes" for some tropical species still lags behind cereals. Fortunately, in the last decades, fast changes have occurred in this scenario, with numerous genomic projects being developed on previously considered less studied crops such as cowpea, chickpea, mung bean and peanut (Varshney et al. 2015; Pandey et al. 2012; Bertioli et al. 2009) that will certainly leverage the understanding of legumes biology, essential to produce new improved varieties.

## 10.2 Transcriptomics in *Arachis*

Transcriptomic analysis has been widely used as an effective and less cost intensive alternative to whole genome sequencing, especially in less studied crops, including peanut. Thus far, most markers development and large-scale discovery, isolation, and deciphering of gene functions in *Arachis* have been acquired by relying on this strategy. Over the years, arrays and sequence-based methods, in special Expressed Sequence Tags (ESTs) studies have generated a large number of cDNA sequences, which provided a considerable amount of transcriptional information on the genus. These studies, which included various peanut genotypes under different stresses and at differential developmental stages were recently revised (Pandey et al. 2012; Feng et al. 2012; Chen and Liang 2014; Varshney et al. 2013). These resources are valuable for the development of novel EST-derived simple sequence repeat (EST-SSR) and Single Nucleotide Polymorphisms (SNPs) markers (Nagy et al. 2012), and also for the identification of genes involved in responses to stress (Brasileiro et al. 2014).

Nevertheless, the generation of ESTs is laborious, requires cloning, and often produces low coverage of less abundant or rare transcripts that might play a vital role in gene modulation. Most of these disadvantages have been overcome by Next Generation Sequencing (NGS) technologies that allow rapid, high quality and massively parallel sequencing of transcript sequences. Given that most NGS methods provide only relatively short sequence read lengths, hindering the production of high-quality de novo assemblies, the employment of different complementary and synergistic strategies are essential to consolidate transcriptome analysis (Chen and Liang 2014).

The arrival of NGS technologies in 2005 has led to a paradigm shift in accessing genetic variations and the role of functional elements in plant biology. The so-called "orphan crops", including peanut, with few genomic assets, have profited enormously from these advances, as the cost of DNA sequencing has dropped by 5 or 10 times per year since the beginning of the millennium (www.genome.gov/sequencingcosts). Moreover, considering the estimate size (2.7 Gb) (Bertioli et al. 2016) and the high repetition content (~64%) (Dhillon et al. 1980) of the peanut genome, high-throughput sequencing (NGS), especially transcriptome sequencing, has shown to be a convenient technique for the study of whole gene expression pattern at a given growth stage or tissue or as a result of a stimulus (Shen et al. 2015) (Table 10.1). In fact, the application of these technologies in peanut, has enabled an unprecedented development of molecular markers, such as SSRs and SNPs

derived from RNA-Seq transcripts, accelerating the saturation and construction of other genetic maps and providing new tools for precision breeding (Varshney et al. 2014; Peng et al. 2015; Chopra et al. 2015; Zhang et al. 2012).

High-throughput transcriptome sequencing was also applied to achieve a high-quality reference transcriptome for three peanut botanical types at different developmental stages (Wu et al. 2013; Chopra et al. 2014; Patel et al. 2015), and the construction of *A. hypogaea* "Tiffrunner" gene atlas comprising 22 different tissues and 260 Gb of data (http://www.peanutbase.org) (Table 10.1). These resources have great potential to be exploited for mapping and QTL analysis, especially considering that peanut still lacks well-annotated genomic resources for genomics-assisted breeding.

RNA-Seq has been applied to decipher peanut genes specifically expressed in some tissues/organs, such as in pods related to early embryo abortion (Chen et al. 2013), to gain insight into the mechanism of geocarpy from gynophores (Xia et al. 2013; Chen et al. 2015b), identify genes involved in lipid metabolism of high- and low-oil varieties (Yin et al. 2013; Yu et al. 2015), and to identify tissue-specific genes for promoter cloning (Geng et al. 2014). Likewise, RNA-Seq helped identify transcripts involved in response to abiotic and biotic stresses in peanut, such as drought (Shen et al. 2015; Li et al. 2014; Chen et al. 2014a) and infection with *Ralstonia solanacearum* (Chen et al. 2014b) (Table 10.1). Deep sequencing of peanut small RNA libraries has also been used to identify large numbers of miRNAs and their related target genes, which provides a greater understanding of the regulation mechanisms and roles of miRNAs in the crop (Zhao et al. 2010; Chi et al. 2011).

Transcriptome profiling has also been extremely valuable for wild *Arachis*, as it offers an unparalleled resolution for interpreting the functional elements of the genome, thus revealing cell responses to stresses. Besides, as this technology is not limited to detecting transcripts that correspond to existing genomic sequences, it has allowed the analysis of poorly characterized *Arachis* wild species with genomic sequences that are yet to be fully determined.

Over the years, a number of reports showed that wild *Arachis* species are resistant to virus (Kalyani et al. 2007; Reddy et al. 2000), foliar and root fungi (Singh et al. 2013), nematodes (Choi et al. 1999; Bendezu and Starr 2003; Proite et al. 2008) and more tolerant than peanut to abiotic stresses, such as drought (Brasileiro et al. 2015; Leal-Bertioli et al. 2012). Nevertheless, breeding efforts to utilize the resistances found in these wild species in order to develop peanut cultivars with multiple resistances, high seed quality, and yield have been hindered by the difference in ploidy and the complex nature of the tetraploid genome, as mentioned above (Pandey et al. 2012; Nagy et al. 2012). The identification of genes and regulators that trigger these resistance responses, their introgression via marker-assisted selection (MAS) or plant genetic transformation constitute, therefore, an interesting alternative for their deployment into high yield cultivars.

In the beginning of 2000's, transcriptomic studies on *Arachis* wild species were limited in magnitude and restricted to few species, constituting mainly of ESTs databases (Table 10.2). Fortunately, the advent of NGS high-throughput sequencing technologies rapidly changed this scenario, enabling the full use of these wild species to contribute to understanding the peanut transcriptome and the global expression variations in response to environmental conditions. In the last five years, the number of genomic resources, including genomic and transcribed sequences, has already increased dramatically for wild species (Table 10.2).

The diploid species *A. duranensis* and *A. ipaënsis* are the parentals of the mapping populations used to develop the reference diploid A and B genetic maps (Moretzsohn et al. 2005, 2009), and for the co-localization of QTLs with candidate genes (Leal-Bertioli et al. 2009). They are also the peanut wild relatives with the greatest arsenal of genomic tools developed so far, such as BAC libraries (Guimaraes et al. 2008) and complete reference genomes sequenced

**Table 10.1** Unique transcripts (Unigenes) in *Arachis hypogaea* generated by different NGS technologies

| *A. hypogaea* | NGS technology | Unigenes | Reference |
|---|---|---|---|
| Oil synthesis | Illumina GA II | 59,236 | Yin et al. (2013) |
| Genotypes | Illumina HiSeq 2000 | 26,048 | Wu et al. (2013) |
| Drought stress | Illumina HiSeq 2000 | 47,842 | Li et al. (2014) |
| Drought stress | Illumina HiSeq 2000 | 62,510 | Shen et al. (2015) |
| Oil synthesis | Illumina GA II | 1500 | Yu et al. (2015) |
| Transcriptome | Illumina HiSeq 2000 | 10,824 | Patel et al. (2015) |
| Gynophores | Illumina HiSeq 2000 | 72,527 | Xia et al. (2013) |
| Embryo abortion | Roche 454 life science | 74,974 | Chen et al. (2013) |
| Geocarpy | Illumina HiSeq 2000 | 110,217 | Chen et al. (2015b) |
| SNP development | Illumina GA II | 43,108 | Chopra et al. (2015) |
| Various tissues | Illumina GA II | 960 | Geng et al. (2014) |
| SSR development | Solexa HiSeq™ 2000 | 59,077 | Zhang et al. (2012) |
| Reference transcriptome | Illumina HiSeq 2000 and Roche 454 life science | 415,942 | Chopra et al. (2014) |
| *Ralstonia solanacearum* | Illumina HiSeq 2000 | 271, 790 | (Chen et al. 2014b) |
| Transcriptome atlas | Illumina HiSeq 2000 | 102,303 | Ozias-Akins et al. (http://peanutbase.org/) |
| SSR development | Roche 454 life science | 44,007 | Peng et al. (2015) |

(http://peanutbase.org/; Bertioli et al. 2016). Likewise, the majority of the transcriptomic studies on wild *Arachis* species, including ESTs and NGS data, include these species (Table 10.2).

One of the main applications of transcriptomic studies in wild *Arachis* species is to circumvent low polymorphisms owing to narrow genetic diversity in the cultivated species. For that, the generation of transcripts from the above species, which are also the most probable peanut diploid parents, were extensively used to facilitate the development of SSRs and SNPs markers. These markers have been useful not only for gene discovery and genetic mapping but also for comparative mapping among diverse populations (Koilkonda et al. 2012; Bertioli et al. 2009; Nagy et al. 2012).

ESTs from *A. duranensis* and *A. ipaënsis* developing seeds tissues were the first available, with some being used for fine mapping of resistance genes in *A. duranensis,* the A diploid progenitor (Nagy et al. 2010). Later, with the application of the 454 sequencing technology, more than one million cDNA sequence reads were generated from different genotypes of *A. duranensis,* which were assembled into 81,116 unique transcripts (Unigenes), and yielded 1236 EST-SNP markers (Nagy et al. 2012).

The application of 454 technology also enabled the generation of 12,792 Unigenes and 1463 EST-SSRs from cDNA libraries of *A. duranensis* roots submitted to gradual water deficit (Guimaraes et al. 2012). This species displayed a conservative transpiration profile under water-limited conditions when compared to other wild and cultivated *Arachis* genotypes, constituting a very interesting species for drought-related gene discovery (Leal-Bertioli et al. 2012). Following studies enabled the identification of a number of candidate genes differentially expressed in water deficit conditions and potentially involved in drought tolerance mechanisms in *A. duranensis* (Brasileiro et al. 2015). Further validation by RT-qPCR revealed differential gene expression modulation of 31 candidates involved in drought perception

**Table 10.2** Unique transcripts (Unigenes) in *Arachis* wild species generated by different sequencing technologies

| Species | Tissue/stress | ESTs | 454 | RNA-Seq | Reference |
|---|---|---|---|---|---|
| *A. duranensis* | Seeds, roots, leaves | 35,291 | – | – | Ozias–Akins et al. (http://peanutbase.org/) |
| *A. ipaënsis* | Roots, developing seeds | – | 81,116 | – | Nagy et al. (2012) |
| | Leaves, roots, pods | – | – | 37,379 | Chopra et al. (2014) |
| | Drought | – | 12,792 | 21,126 | Guimaraes et al. (2012) Brasileiro et al. (http://peanutbase.org/ |
| | *Meloidogyne arenaria* | – | – | 25,844 | Guimarães et al. (http://peanutbase.org/) |
| | Seeds, roots, leaves | 32,787 | – | – | Nagy et al. (http://www.ncbi.nlm.nih.gov/) |
| *A. stenosperma* | Leaves, roots, pods | – | – | 31,800 | Guimaraes et al. (2012) |
| | *Meloidogyne arenaria* | 6264 | – | 44,132 | Proite et al. (2007), Guimaraes et al. (2015) |
| *A. magna* | *Phaeoisariopsis personata* | – | 7723 | – | Guimaraes et al. (2012) |
| | Drought | – | – | 21,503 | Brasileiro et al. (http://peanutbase.org/) |
| | Drought | 750 | – | – | Brasileiro et al. (2015) |
| *A. diogoi* | *Phaeoisariopsis personata* | 308 | – | – | Kumar and Kirti (2011, 2015) |
| *A. appressipila* | *Ralstonia solanacearum* | 400 | – | – | Chen et al. (www.ncbi.nlm.nih.gov) |

and key processes underlying plant tolerance to water-limited conditions, such as Expansins, Late Embryogenesis Abundant (LEA), Nitrilase and P450 proteins and NAC, bZIP, ERF and MYB transcription factors (Brasileiro et al. 2015).

More recently, RNA-Seq technology was applied aiming to produce a good quality reference transcriptome for *A. ipaënsis* and *A. duranensis* using different de novo assembling for both tetraploid and diploid species (Chopra et al. 2014). These assemblies from leaves, roots and different maturation stages of pods of *A. ipaënsis* and *A. duranensis* comprised 31,800 and 37,379 Unigenes respectively, which will be useful for SNP selection, expression analysis, mapping and QTL analysis.

Another wild species, *A. stenosperma,* shows resistances to a number of biotic stresses, including joint foliar diseases (early and late leaf spot, rust and scab), constituting one of the most resistant genotypes to be used for multiple selections and a candidate species for gene discovery (Nelson et al. 1989; Singsit et al. 1995; Michelotto et al. 2015; Leal-Bertioli et al. 2010). Therefore, 454 technology was applied for the elucidation of *A. stenosperma* genes involved in the high levels of resistance to late leaf spot (*Phaeoisariopsis personata*) and for molecular markers development (Guimaraes et al. 2012). High-quality reads generated a total of 7,723 Unigenes from *A. stenosperma* leaves infected with *P. personata,* enabling the identification of 20 transcription factor gene families, defense-related genes, including Resistance Gene Analogs (RGAs), and retrotransposon (FIDEL-related) sequences (Guimaraes et al. 2012). In this work, another set of 862 EST-SSRs was also designed, which contributed to the enrichment of the A genome diploid map, and the identification of QTLs related to disease resistances (Leal-Bertioli et al. 2009; Shirasawa et al. 2013).

*A. stenosperma* has also been identified as harboring high levels of resistance to the peanut RKN *M. arenaria,* through the onset of the hypersensitive response (HR) (Proite et al. 2008). Transcriptome studies aiming the identification of genes differentially expressed between nematode infected and control roots of *A. stenoperma* identified genes related to HR response and produced the first ESTs collection and SSR-ESTs derived for the species (Proite et al. 2007). Candidate genes from these previous studies involved in HR, hormonal balance, production of secondary metabolites or related to pathogen defense, such as resistance protein MG13, auxin-repressed protein (ARP), metallothionenin, patatin-like protein, catalase, DUF protein, and resveratrol synthase had their expression profiles further validated by macroarray and RT-qPCR (Guimarães et al. 2010; Morgante et al. 2013). Later on, RNA–Seq was applied to produce a comprehensive root transcriptome on the first stages of this incompatible interaction, and to identify specific genes and regulators triggering the HR defense response. Overall, 28.2 Gb of transcript data was generated and de novo assembled into 44,132 Unigenes. Differentially expressed genes were identified and further validated by RT-qPCR, allowing the identification of nematode-responsive candidate genes engaged in the salicylic (NBS-LRR, lipocalins, resveratrol synthase) and jasmonic (patatin, allene oxidase cyclase) acids pathways, and also related to hormonal balance (auxin responsive protein, GH3) and cellular plasticity and signaling (tetraspanin, integrin, expansin). Some of these genes showed contrasting expression behavior between *Arachis* RKN-resistant and susceptible genotypes (Guimaraes et al. 2015).

Unlike *A. duranensis* and *A. ipaënsis*, and to a lesser extent *A. stenosperma*, the remaining *Arachis* wild species have been virtually unexploited in terms of their potential as resistance allele donors, with few ESTs publicly available (Table 10.2). The solo study on *A. magna* transcriptome aimed at drought-related gene discovery, as this species displays one of the most conservative transpiration profiles under water-limited conditions when compared to other *Arachis* genotypes (Leal-Bertioli et al. 2012). Leaves and roots of this B genome representative were used to construct suppression subtractive hybridization (SSH) cDNA libraries from plants submitted to progressive water deficit in soil that revealed 757 high-quality ESTs clustered into 284 Unigenes (Brasileiro et al. 2015). Among these, 13 differentially expressed candidate genes were selected for further RT-qPCR validation, allowing the identification of genes associated with the response to water deprivation and to signal transduction pathways under drought, such as carbonic anhydrase, metallothionein, drought-induced proteins (DiP), Ca2 +/H + exchanger, aminomethyltransferase and chloroplast drought-induced stress Protein (CDSP) and (Brasileiro et al. 2015). This study was the unique report on the identification of differentially expressed genes for an *Arachis* B genome species.

The transcripts profiles of the highly resistant *A. diogoi* upon treatment with *P. personata* were analyzed using Differential Display and cDNA-AFLP approaches, and several genes significantly upregulated and related to defense response were identified (Kumar and Kirti 2015, 2011). Some of these genes showed enhanced anti-fungal activity against different pathogens, and also enhanced tolerance against salt and oxidative stress in tobacco transgenic plants (Singh et al. 2013). In the section Procumbentes, *A. appressipila,* known for its resistance to the Peanut Bud Necrosis Virus (PBNV) (Reddy et al. 2000), was the first species to have ESTs publicly available from a transcriptome study using SSH libraries constructed with transcripts in response to the bacterial pathogen *Ralstonia solanacearum* (Table 10.2).

Resistance alleles from wild relatives have already been used for peanut breeding for pest and disease resistances, and there are strong reasons to believe that introducing allelic diversity from wild species could also improve more complex traits such as yield and drought responses (Leal-Bertioli et al. 2015). The availability of reference genomes for the peanut

parents (http://www.peanutbase.org), a fully sequenced genome for the tetraploid cultivated species, which is now on the verge of being accomplished by the International Peanut Genome Initiative (IPGI) (http://www.peanutbioscience.com/homepage.html), together with high-throughput technology for the analyses of transcripts, proteins, and mutants, will contribute to provide the basis for understanding the relationships between genes, proteins and phenotypes.

## 10.3 Proteomics Studies of *Arachis*

Proteomics is a powerful approach that involves the analysis of the entire group of proteins in a specific biological condition. The study of the proteome of an organism is of extreme importance since the final product of gene expression is assessed leading to a better understanding of the phenotype. Several proteomic studies have been performed in peanut, mainly to analyze three major aspects, which are seed storage proteins (Viquez et al. 2003; Kang and Gallo 2007; Koppelman et al. 2004), development (Zhu et al. 2013; Zhao et al. 2015; Sun et al. 2013) and allergen proteins (Kottapalli et al. 2008; Schmidt et al. 2009). However, few studies on the effects of biotic and abiotic stresses at the protein level have been reported for the entire genus (Katam et al. 2016; Wang et al. 2010; Kottapalli et al. 2013; Wang et al. 2012).

Most of the proteomic analysis in *Arachis* spp. has been performed using the classical 2-DE approach, which involves the separation of proteins using two-dimensional gel electrophoresis (Schmidt et al. 2009; Sun et al. 2013; Zhu et al. 2013). Advances in mass spectrometry (MS) over the last years have brought new alternatives for high-throughput proteomic analyses, which rely on liquid chromatographic methods for protein separation and further identification by mass spectrometry (LC-MS). This approach was employed to analyze peanut gynophores during development, with more than 2000 proteins being identified, demonstrating the capacity of MS-based methods to identify a high number of proteins in a relatively short time (Zhao et al. 2015).

One single study has been conducted in wild *Arachis* comprising the analysis of the proteome of the highly resistant *A. diogoi* in response to the fungus *Phaeoisariopsis personata* (Kumar and Kirti 2015). Several proteins involved in defense and signal transduction were identified giving new insights into this plant-fungus interaction. Similarly, the HR response of *A. stenosperma* to *M. arenaria* infection has been analyzed by the 2-DE approach, with several differentially abundant proteins potentially involved in this resistance identified (unpublished data) and correspondent to some genes found to be differentially expressed in previous transcriptome RNA–Seq analysis (Guimaraes et al. 2015).

With respect to abiotic stresses, proteins related to drought tolerance in cultivated peanut have already been identified, such as Lipoxygenase and 1 l-myo-inositol-1-phosphate synthase (Kottapalli et al. 2009) and a leaf protein reference map of a highly drought-tolerant peanut cultivar (Vemana) was developed (Katam et al. 2010), but no reports on wild *Arachis* have yet been released. Recently, *A. duranensis* responses to gradual water deficit in soil have been analyzed by 2-DE and RNA-Seq approaches, with several genes and correspondent proteins that are potentially involved in drought tolerance being differentially down- or upregulated (unpublished data).

Considering the limited proteomics data available on *Arachis* species, efforts are underway for the identification of proteins specific to tissues, developmental stages and responsive to biotic and abiotic stresses. The use of gel-free comprehensive proteome analysis constitutes an important advance which will enable gathering more proteomic information which, coupled to systematic transcriptome surveys, may reveal relevant insights in coordinated transcriptional and posttranscriptional events involved in resistance responses to stresses in wilds *Arachis* species.

## 10.4 Genetic Transformation in *Arachis*

Due to the enormous amount of information generated by different "omics" projects, functional characterization of novel candidate genes has become the main bottleneck to the challenge of associating genes with phenotypes. In the process to validate and associate a biological function to these genes, a critical step is the analysis of the pleiotropic effects and phenotypes associated with their introduction and expression in transgenic plants. This includes candidate gene overexpression and *knockout*; analysis of promoter activity; transient gene expression using strategies like VIGS (Virus Induced Gene Silencing) and root transformation mediated by *Agrobacterium rhizogenes*. These methods allow the assessment of large numbers of candidate genes before its introduction in the target plant, which is, in general, a long, laborious and expensive process, mostly with plants recalcitrant to genetic transformation, such as peanut.

### 10.4.1 Functional Validation in Planta

Studies involving the ectopic expression of transgenes isolated from peanut in *Arabidopsis* have provided useful information for elucidating their biological functions and unraveling their potential role in the control of traits (Liu et al. 2011; Wan and Li 2006; Li et al. 2013; Chen et al. 2015a). The phenotype-associated effects of these peanut transgenes included the improvement for drought and salt tolerance. *Arabidopsis* has also been exploited as a model to analyze the regulation and activity of promoter regions of peanut genes, in particular, tissue- or temporal-specific, and stress-responsive promoters (Geng et al. 2014; Liang et al. 2009; Zhou et al. 2014; Bhattacharya et al. 2012). Such studies are strategic to uncover novel promoters as an alternative to those traditionally used in peanut transformation (35SCaMV, ubiquitin, and actin) (Porto et al. 2014), which have commonly been associated with undesirable side effects in transgenic plants under field conditions. These alternative peanut promoters are thus valuable tools for driving specific gene expression in upcoming genetic and metabolic engineering approaches.

In addition, tobacco (*Nicotiana tabacum*) have also been employed as a model system to validate peanut candidate genes and to untangle their functions and biotechnological applications, particularly those related to biotic and abiotic stresses (Kellmann et al. 1996; Liu et al. 2013; Jain et al. 2012). Recently, three defense-related genes induced in *A. diogoi* infected with *P. personata* (Kumar and Kirti 2011) were overexpressed in transgenic tobacco. The first encoded a thaumatin-like protein, which enhanced the resistance to the fungal pathogen *Rhizoctonia solani*, and tolerance to salinity and oxidative stress (Singh et al. 2013). In the same way, the other genes, one coding for a suppressor of SKP1 (AdSGT1) and the other for a vacuolar processing enzyme (AdVPE), increased the resistance against other plant fungal pathogens (*Phytophthora parasitica, Alternaria alternata* and *R. solani*) (Kumar and Kirti 2015) (Kumar et al. 2015). To our knowledge, these are the few examples of wild *Arachis* genes expression in transgenic model plants targeting biotic and abiotic stress resistances. Moreover, another gene from *A. diogoi* coding for a cysteine protease (AdCP) was used to induce pollen abortion and male sterility in transgenic tobacco (Shukla et al. 2014) using a strong tapetum-specific promoter.

In addition to gene overexpression, transient induction of gene knockdown is also useful for candidate gene validation. Gene silencing induced by viruses (VIGS) is a rapid and simple alternative method to study the effects of down-regulation of candidate genes, mostly in species recalcitrant to stable genetic transformation. By transient gene knockdown induction, VIGS enables analysis of a large set of genes in a relatively short time, without the need for production of stable transgenic plants (Robertson 2004). VIGS was successfully used to validate peanut candidate genes responsive to drought stress, by silencing their corresponding orthologous in *N. benthamiana* (Govind et al. 2009; Senthil-Kumar

et al. 2007), thus corroborating it as a valuable tool for characterization and assessment of *Arachis* spp. genes role in heterologous systems.

Morphologically altered roots, known as hairy roots, are induced by *A. rhyzogenes* and enable the expression and production in a large scale of important secondary metabolites, as resveratrol and other stilbenoids in peanut reviewed by (Hasan et al. 2013). Peanut hairy roots have emerged as an alternative strategy to evaluate the effect of overexpression or silencing of candidate genes, and also to analyze responses to external stimuli, such as interaction with parasites and symbionts. A number of candidate genes have been evaluated in peanut composite plants, in particular, those involved in nodule formation and nitrogen fixation (Akasaka et al. 1998; Sinharoy et al. 2009), interaction with nematode (Chu et al. 2014) and subterranean insects (Geng et al. 2012).

Due to the efficiency and convenience of obtaining *A. rhizogenes*-derived composite plants, it became the large-scale method of choice for in-root screening of candidate genes in many target plant species. For peanut, our group has recently established a faster and simpler method to obtain composite plants for in-root functional characterization of candidate genes. The premise is to take advantage of the well-known ability of *Arachis* spp. to naturally produce roots originating from the petiole of detached leaves (Fávero et al. 2004; Melouk and Banks 1978; Subrahmanyam et al. 1983), which can be cultured and maintained under moistened conditions for extended periods. Such habit has widely being used for swift phenotyping of *Arachis* genotypes for resistance to foliar diseases and insect pests (Leal-Bertioli et al. 2009; Sharma et al. 2005; Subrahmanyam et al. 1983). This new method takes advantage of this rooting competence to accelerate candidate genes functional analysis in peanut cultivars. In short, the youngest fully expanded quadrifoliate leaves are harvested from six-week-old peanut plants and placed on sterile Petri dishes with the adaxial surface side up (Fig. 10.1), as previously described (Leal-Bertioli et al. 2009). The petiole base of detached leaves is immediately inoculated with *A. rhizogenes* K599 wild-type strain harboring the pPZP-201BK-eGFP binary vector (Chu et al. 2014) and covered with moistened sterile vermiculite on Petri dishes. Inoculated leaves were maintained at 23-25 °C, under a 16 h photoperiod. High transformation efficiency was obtained 30 days after inoculation, when 95% of emerged hairy roots were GFP-positive (Fig. 10.1). The use of inoculated detached leaves for hairy root production in peanut revealed to be a rapid, simple and efficient method for screening in-root expression of candidate genes. This method is currently being optimized for soybean and considered for other species, such as chickpea, pigeon pea, and cotton, in which development of roots originating from the petiole of detached leaves has been successfully tested (Sharma et al. 2005; Fávero et al. 2004).

### 10.4.2 Stable Genetic Transformation

The first transgenic peanut plant was achieved in 1993 (Ozias-Akins et al. 1993), and since then, various protocols to obtain stable and fertile transgenic peanut plants have been described, which rely on efficient whole plant regeneration systems. Nevertheless, transformation protocols mediated by *Agrobacterium* and biolistics showed, in general, genotype dependency, low regeneration frequencies and decreased transformation efficiency (Ozias-Akins and Gill 2001; Chu et al. 2013; Holbrook et al. 2011).

Several genes have been already introduced and stably expressed in different peanut cultivars, resulting in the improvement of agronomical traits such as resistance/tolerance to biotic and abiotic stresses; herbicide resistance and seed quality, and for increased production of biopharmaceuticals. These studies were reviewed recently, highlighting the scarcity of transgenes originated from *Arachis* species (Krishna et al. 2015; Holbrook et al. 2011; Brasileiro et al. 2014). It is anticipated that with the recent availability of a great number of genome and transcriptome sequences from wild and cultivated *Arachis,* a much broader collection of

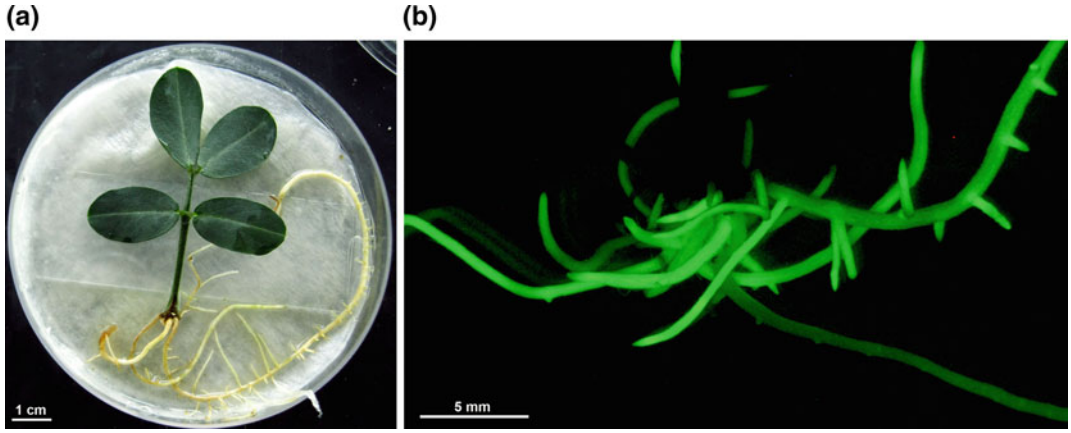


**Fig. 10.1** Detached leaf of peanut (*Arachis hypogaea*) 'IAC—Runner' after 30 days of inoculation with *Agrobacterium rhyzogenes* K599/pPZP-201BK-EGFP showing the hairy root phenotype (**a**) with fluorescent eGFP constitutive expression (**b**) *Credits* to Larissa Arrais Guimaraes and Bruna Medeiros Pereira

candidate genes will become available in relatively short time.

Whereas the ultimate validation of candidate genes in transgenic targeted-plants is not an easy task for those, which are recalcitrant to genetic transformation, such as peanut, only the genes with a validated function in model systems will be select for further steps. Once produced, promising transgenic events might be directly used as elite-events, or as donors in breeding programs aiming improved cultivars. In addition, validated genes can be mapped and converted into functional molecular markers that co-segregate with traits of interest and be used in MAS in peanut breeding programs.

## 10.5 Conclusion

Rapid developments in NGS technologies have allowed the sequencing of two wild *Arachis* genomes and few transcriptomes of wild species under biotic and abiotic stresses. Nevertheless, considering the wealth and diversity of the species in the genus in terms of tolerance to environmental stresses and resistance to parasites and pests, much more is yet to be explored to unlock the genetic diversity available in germplasm banks. In addition, with greater resolution, genes and pathways triggered in responses to stresses in *Arachis* are becoming unraveled, thus clarifying the relationships between genotype/phenotype. Likewise, recent progress in proteomics have the potential to uncover the roles of ultimate genes products in *Arachis*, changed by physiological processes, environment, and pathogens.

Nevertheless, in order to fully profit from the increasingly amount of *Arachis* "omics" resources for peanut breeding, more efficient and integrated data analysis and large-scale phenotyping are required. As so, advances in transient and stable peanut transformation protocols and development of scalable methods for candidate genes validation should be urgently incremented. These improvements will certainly contribute to overcoming the crossing barriers between peanut and its wild relatives that for so long have hampered the effective use of wild alleles in peanut improvement.

## References

Abbo S, Lev-Yadun S, Gopher A (2012) Plant domestication and crop evolution in the Near East: on events and processes. Crit Rev Plant Sci 31(3):241–257

Akasaka Y, Mii M, Daimon H (1998) Morphological alterations and root nodule formation in

*Agrobacterium rhizogenes*-mediated transgenic hairy roots of peanut (*Arachis hypogaea* L.). Ann Bot 81 (2):355–362

Bendezu IF, Starr JL (2003) Mechanism of Resistance to *Meloidogyne arenaria* in the peanut cultivar COAN. J Nematology 35(1):115–118

Bertioli DJ, Moretzsohn MC, Madsen LH, Sandal N, Leal-Bertioli SC, Guimarães PM, Hougaard BK, Fredslund J, Schauser L, Nielsen AM (2009) An analysis of synteny of *Arachis* with *Lotus* and *Medicag*o sheds new light on the structure, stability and evolution of legume genomes. BMC Genom 10 (1):45

Bertioli DJ, Seijo G, Freitas FO, Valls JF, Leal-Bertioli S, Moretzsohn MC (2014) An overview of peanut and its wild relatives. Plant Genet Resour 9(01):134–149

Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EKS, Liu X, Gao D, Clevenger J, Dash S, Ren L, Moretzsohn MC, Shirasawa K, Huang W, Vidigal B, Abernathy B, Chu Y, Niederhuth CE, Umale P, Araujo ACG, Kozik A, Do Kim K, Burow MD, Varshney RK, Wang X, Zhang X, Barkley N, Guimaraes PM, Isobe S, Guo B, Liao B, Stalker HT, Schmitz RJ, Scheffler BE, Leal-Bertioli SCM, Xun X, Jackson SA, Michelmore R, Ozias-Akins P (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet Adv Online Publ 48:438–446

Bhattacharya A, Ramos M, Faustinelli P, Ozias-Akins P (2012) Reporter gene expression patterns regulated by an Ara h 2 promoter differ in homologous versus heterologous systems. Peanut Sci 39(1):43–52

Bonavia D (1982) Los Gavilanes: mar, desierto y oásis en la historia del hombre: precerámico peruano. Corporación Financiera de Desarrollo, Oficina de Asuntos Culturales

Brasileiro ACM, Araujo ACG, Leal-Bertioli SC, Guimaraes PM (2014) Genomics and genetic transformation in *Arachis*. Int J Plant Biol Res 2(3):1017

Brasileiro AM, Morgante C, Araujo AG, Leal-Bertioli SM, Silva A, Martins AQ, Vinson C, Santos CR, Bonfim O, Togawa R, Saraiva MP, Bertioli D, Guimaraes P (2015) Transcriptome profiling of wild *Arachis* from water-limited environments uncovers drought tolerance candidate genes. Plant Mol Biol Rep:1–17

Brozynska M, Furtado A, Henry RJ (2015) Genomics of crop wild relatives: expanding the gene pool for crop improvement. Plant Biotechnol J

Chen X, Liang X (2014) Peanut transcriptomics. Genet, Genomics and Breeding of Peanuts:139

Chen X, Zhu W, Azam S, Li H, Zhu F, Li H, Hong Y, Liu H, Zhang E, Wu H, Yu S, Zhou G, Li S, Zhong N, Wen S, Li X, Knapp SJ, Ozias-Akins P, Varshney RK, Liang X (2013) Deep sequencing analysis of the transcriptomes of peanut aerial and subterranean young pods identifies candidate genes related to early embryo abortion. Plant Biotechnol J 11(1):115–127

Chen N, Yang Q, Pan L, Chi X, Chen M, Hu D, Yang Z, Wang T, Wang M, Yu S (2014a) Identification of 30 MYB transcription factor genes and analysis of their expression during abiotic stress in peanut (*Arachis hypogaea* L.). Gene 533(1):332–345

Chen Y, Ren X, Zhou X, Huang L, Yan L, Lei Y, Liao B, Huang J, Huang S, Wei W, Jiang H (2014b) Dynamics in the resistant and susceptible peanut (*Arachis hypogaea* L.) root transcriptome on infection with the *Ralstonia solanacearum*. BMC Genomics 15 (1):1078

Chen S, Lei Y, Xu X, Huang J, Jiang H, Wang J, Cheng Z, Zhang J, Song Y, Liao B (2015a) The peanut (*Arachis hypogaea* L.) gene AhLPAT2 Increases the lipid content of transgenic *Arabidopsis* seeds. PLoS ONE 10(8):e0136170

Chen X, Yang Q, Li H, Li H, Hong Y, Pan L, Chen N, Zhu F, Chi X, Zhu W, Chen M, Liu H, Yang Z, Zhang E, Wang T, Zhong N, Wang M, Liu H, Wen S, Li X, Zhou G, Li S, Wu H, Varshney R, Liang X, Yu S (2015b) Transcriptome-wide sequencing provides insights into geocarpy in peanut (*Arachis hypogaea* L.). Plant Biotechnol J 14(5):1214–1224

Chi X, Yang Q, Chen X, Wang J, Pan L, Chen M, Yang Z, He Y, Liang X, Yu S (2011) Identification and characterization of microRNAs from peanut (*Arachis hypogaea* L.) by high-throughput sequencing. PLoS ONE 6(11):e27530

Choi K, Burow MD, Church G, Burow G, Paterson AH, Simpson CE, Starr JL (1999) Genetics and mechanism of resistance to *Meloidogyne arenaria* in peanut germplasm. J Nematology 31(3):283–290

Chopra R, Burow G, Farmer A, Mudge J, Simpson CE, Burow MD (2014) Comparisons of *de novo* transcriptome assemblers in diploid and polyploid species using peanut (*Arachis* spp.) RNA-Seq data. PLoS ONE 9(12):e115055

Chopra R, Burow G, Farmer A, Mudge J, Simpson C, Wilkins T, Baring M, Puppala N, Chamberlin K, Burow M (2015) Next-generation transcriptome sequencing, SNP discovery and validation in four market classes of peanut. Arachis hypogaea L. Mol Genet Genomics 290(3):1169–1180

Chu Y, Bhattacharya A, Wu C, Knoll J, Ozias-Akins P (2013) Improvement of peanut (*Arachis hypogaea* L.) transformation efficiency and determination of transgene copy number by relative quantitative real-time PCR. In Vitro Cell Dev Biol -Plant 49(3):266–275

Chu Y, Guimaraes L, Wu C, Timper P, Holbrook C, Ozias-Akins P (2014) A technique to study *Meloidogyne arenaria* resistance in *Agrobacterium rhizogenes*-transformed peanut. Plant Dis 98(10):1292–1299

Devi M, Bhatnagar-Mathur P, Sharma K, Serraj R, Anwar S, Vadez V (2011) Relationships between transpiration efficiency and its surrogate traits in the rd29A:DREB1A transgenic lines of groundnut. J Agronomical Crop Sci 197:272–283

Dhillon SS, Rake AV, Miksche JP (1980) Reassociation kinetics and cytophotometric characterization of

peanut (*Arachis hypogaea* L.) DNA. Plant Physiol 65 (6):1121–1127

Dwivedi S, Crouch J, Nigam S, Ferguson M, Paterson A (2003) Molecular breeding of groundnut for enhanced productivity and food security in the semi-arid tropics: opportunities and challenges. Adv Agron 80:153–221

Fávero AP, Cuco SM, Aguiar-Perecin MLd, Valls JF, Vello NA (2004) Rooting in leaf petioles of *Arachis* for cytological analysis. Cytologia 69(2):215–219

Feng S, Wang X, Zhang X, Dang PM, Holbrook CC, Culbreath AK, Wu Y, Guo B (2012) Peanut (*Arachis hypogaea*) expressed sequence tag project: progress and application. Comp Funct Genomics 2012:9

Geng L, Niu L, Gresshoff P, Shu C, Song F, Huang D, Zhang J (2012) Efficient production of Agrobacterium rhizogenes-transformed roots and composite plants in peanut (*Arachis hypogaea* L.). Plant Cell, Tissue and Organ Culture (PCTOC) 109(3):491–500

Geng L, Duan X, Liang C, Shu C, Song F, Zhang J (2014) Mining tissue-specific contigs from peanut (*Arachis hypogaea* L.) for promoter cloning by deep transcriptome sequencing. Plant Cell Physiol 55(10):1793–1801

Gepts P (2014) The contribution of genetic and genomic approaches to plant domestication studies. Curr Opin Plant Biol 18:51–59

Govind G, Vokkaliga ThammeGowda H, Jayaker Kalaiarasi P, Iyer D, Muthappa S, Nese S, Makarla U (2009) Identification and functional validation of a unique set of drought induced genes preferentially expressed in response to gradual water stress in peanut. Mol Genet Genomics 281(6):591–605

Guimaraes P, Garsmeur O, Proite K, Leal-Bertioli S, Seijo G, Chaine C, Bertioli D, D'Hont A (2008) BAC libraries construction from the ancestral diploid genomes of the allotetraploid cultivated peanut. BMC Plant Biol 8(1):14

Guimaraes P, Brasileiro A, Morgante C, Martins A, Pappas G, Silva O, Togawa R, Leal-Bertioli S, Araujo A, Moretzsohn M, Bertioli D (2012) Global transcriptome analysis of two wild relatives of peanut under drought and fungi infection. BMC Genom 13 (1):387

Guimaraes PM, Guimaraes LA, Morgante CV, Silva OB Jr, Araujo ACG, Martins ACQ, Saraiva MAP, Oliveira TN, Togawa RC, Leal-Bertioli SCM, Bertioli DJ, Brasileiro ACM (2015) Root transcriptome analysis of wild peanut reveals candidate genes for nematode resistance. PLoS ONE 10(10):e0140937

Guimarães P, Brasileiro A, Proite K, de Araújo A, Leal-Bertioli S, Pic-Taylor A, da Silva F, Morgante C, Ribeiro S, Bertioli D (2010) A study of gene expression in the nematode resistant wild peanut relative, *Arachis stenosperma,* in response to challenge with *Meloidogyne arenaria*. Tropical Plant Biol:1–10

Hasan MM, Cha M, Bajpai VK, Baek K-H (2013) Production of a major stilbene phytoalexin, resveratrol in peanut (*Arachis hypogaea*) and peanut products: a mini review. Rev Environ Sci Bio/Technol 12(3):209–221

Holbrook C, Stalker H (2003) Peanut breeding and genetic resources. Plant Breeding Rev 22:297–355

Holbrook CC, Ozias-Akins P, Chu Y, Guo B (2011) Impact of molecular genetic research on peanut cultivar development. Agronomy 1(1):3–17

Husted L (1936) Cytological studies an the peanut. Arachis. II. Cytologia 7(3):396–423

Jain S, Kumar D, Jain M, Chaudhary P, Deswal R, Sarin NB (2012) Ectopic overexpression of a salt stress-induced pathogenesis-related class 10 protein (PR10) gene from peanut (Arachis hypogaea L.) affords broad spectrum abiotic stress tolerance in transgenic tobacco. Plant Cell, Tissue and Organ Culture (PCTOC) 109(1):19–31

Janila P, Nigam SN, Pandey MK, Nagesh P, Varshney RK (2013) Groundnut improvement: use of genetic and genomic tools. Frontiers in Plant Sci 4:23

Kalyani G, Reddy AS, Kumar PL, Rao RDVJP, Aruna R, Waliyar F, Nigam SN (2007) Sources of resistance to tobacco streak virus in wild *Arachis* (Fabaceae: Papilionoidae) Germplasm. Plant Dis 91(12):1585–1590

Kang IH, Gallo M (2007) Cloning and characterization of a novel peanut allergen Ara h 3 isoform displaying potentially decreased allergenicity. Plant Sci 172 (2):345–353

Katam R, Basha SM, Suravajhala P, Pechan T (2010) Analysis of peanut leaf proteome. J Proteome Res 9 (5):2236–2254

Katam R, Sakata K, Suravajhala P, Pechan T, Kambiranda DM, Naik KS, Guo B, Basha SM (2016) Comparative leaf proteomics of drought-tolerant and-susceptible peanut in response to water stress. J Proteomics 143(30):209–226

Kellmann J-W, Kleinow T, Engelhardt K, Philipp C, Wegener D, Schell J, Schreier PH (1996) Characterization of two class II chitinase genes from peanut and expression studies in transgenic tobacco plants. Plant Mol Biol 30(2):351–358

Kochert G, Stalker HT, Gimenes M, Galgaro L, Lopes CR, Moore K (1996) RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). Am J Bot:1282–1291

Koilkonda P, Sato S, Tabata S, Shirasawa K, Hirakawa H, Sakai H, Sasamoto S, Watanabe A, Wada T, Kishida Y, Tsuruoka H, Fujishiro T, Yamada M, Kohara M, Suzuki S, Hasegawa M, Kiyoshima H, Isobe S (2012) Large-scale development of expressed sequence tag-derived simple sequence repeat markers and diversity analysis in *Arachis* spp. Mol Breeding 30(1):125–138

Koppelman SJ, Wensing M, Ertmann M, Knulst AC, Knol EF (2004) Relevance of Ara h1, Ara h2 and Ara h3 in peanut-allergic patients, as determined by immunoglobulin E Western blotting, basophil–histamine release and intracutaneous testing: Ara h2 is

the most important peanut allergen. Clin Exp Allergy 34(4):583–590

Kottapalli KR, Payton P, Rakwal R, Agrawal GK, Shibato J, Burow M, Puppala N (2008) Proteomics analysis of mature seed of four peanut cultivars using two-dimensional gel electrophoresis reveals distinct differential expression of storage, anti-nutritional, and allergenic proteins. Plant Sci 175(3):321–329

Kottapalli KR, Rakwal R, Shibato J, Burow G, Tissue D, Burke J, Puppala N, Burow M, Payton P (2009) Physiology and proteomics of the water-deficit stress response in three contrasting peanut genotypes. Plant, Cell Environ 32(4):380–407

Kottapalli KR, Zabet-Moghaddam M, Rowland D, Faircloth W, Mirzaei M, Haynes PA, Payton P (2013) Shotgun label-free quantitative proteomics of water-deficit-stressed midmature peanut (*Arachis hypogaea* L.) seed. J Proteome Res 12(11):5048–5057

Krapovickas A, Gregory W.C (1994) Taxonomia del Gênero Arachis (Leguminosae). Bonplandia 8:1–186

Krapovickas A, Gregory WC, Williams DE, Simpson CE (2007) Taxonomy of the genus *Arachis* (LEGUMINOSAE). Bonplandia 16:7–205

Krishna G, Singh BK, Kim EK, Morya VK, Ramteke PW (2015) Progress in genetic engineering of peanut (*Arachis hypogaea* L.)—A review. Plant Biotechnol J 13(2):147–162

Kumar K, Kirti P (2011) Differential gene expression in *Arachis diogoi* upon interaction with peanut late leaf spot pathogen, *Phaeoisariopsis personata* and characterization of a pathogen induced cyclophilin. Plant Mol Biol 75(4):497–513

Kumar D, Kirti PB (2015) Transcriptomic and proteomic analyses of resistant host responses in *Arachis diogoi* challenged with Late Leaf Spot pathogen. Phaeoisariopsis personata. PLoS ONE 10(2):e0117559

Kumar D, Rampuria S, Singh N, Shukla P, Kirti PB (2015) Characterization of a vacuolar processing enzyme expressed in *Arachis diogoi* in resistance responses against Late Leaf Spot pathogen. Phaeoisariopsis personata. Plant Mol Biol 88(1–2):177–191

Leal-Bertioli S, Jose A, Alves-Freitas D, Moretzsohn M, Guimaraes P, Nielen S, Vidigal B, Pereira R, Pike J, Favero A, Parniske M, Varshney R, Bertioli D (2009) Identification of candidate genome regions controlling disease resistance in *Arachis*. BMC Plant Biol 9 (1):112

Leal-Bertioli SC, De Farias MP, Silva Pedro ÍT, Guimarães PM, Brasileiro ACM, Bertioli DJ, De Araujo ACG (2010) Ultrastructure of the initial interaction of *Puccinia arachidis* and *Cercosporidium personatum* with leaves of *Arachis hypogaea* and *Arachis stenosperma*. J Phytopathol 158(11–12):792–796

Leal-Bertioli SC, Bertioli DJ, Guimarães PM, Pereira TD, Galhardo I, Silva JP, Brasileiro ACM, Oliveira RS, Silva PÍ, Vadez V (2012) The effect of tetraploidization of wild *Arachis* on leaf morphology and other drought-related traits. Environ Exp Bot 84:17–24

Leal-Bertioli SCM, Santos SP, Dantas KM, Inglis PW, Nielen S, Araujo ACG, Silva JP, Cavalcante U, Guimarães PM, Brasileiro ACM, Carrasquilla-Garcia N, Penmetsa RV, Cook D, Moretzsohn MC, Bertioli DJ (2015) *Arachis batizocoi*: a study of its relationship to cultivated peanut (*A. hypogaea*) and its potential for introgression of wild genes into the peanut crop using induced allotetraploids. Ann Bot 115(2):237–249

Li X-Y, Liu X, Yao Y, Li Y-H, Liu S, He C-Y, Li J-M, Lin Y-Y, Li L (2013) Overexpression of *Arachis hypogaea* AREB1 gene enhances drought tolerance by modulating ROS scavenging and maintaining endogenous ABA content. Int J Mol Sci 14(6):12827–12842

Li X, Lu J, Liu S, Liu X, Lin Y, Li L (2014) Identification of rapidly induced genes in the response of peanut (*Arachis hypogaea*) to water deficit and abscisic acid. BMC Biotechnol 14(1):58

Liang J, Yang L, Chen X, Li L, Guo D, Li H, Zhang B (2009) Cloning and characterization of the promoter of the 9-cis-epoxycarotenoid dioxygenase gene in *Arachis hypogaea* L. Biosci Biotech Bioch 73 (9):2103–2106

Liu X, Hong L, Li X-Y, Yao Y, Hu B, Li L (2011) Improved drought and salt tolerance in transgenic *Arabidopsis* overexpressing a NAC transcriptional factor from *Arachis hypogaea*. Biosci Biotech Bioch 75(3):443–450

Liu X, Liu S, Wu J, Zhang B, Li X, Yan Y, Li L (2013) Overexpression of Arachis hypogaea NAC3 in tobacco enhances dehydration and drought tolerance by increasing superoxide scavenging. Plant Physiol Biochem 70:354–359

Melouk H, Banks D (1978) A method of screening peanut genotypes for resistance to Cercospora leafspot 1. Peanut Sci 5(2):112–114

Michelotto MD, Barioni W, Jr., de Resende MDV, de Godoy IJ, Leonardecz E, Fávero AP (2015) Identification of fungus resistant wild accessions and interspecific hybrids of the genus *Arachis*. PLoS ONE 10 (6):e0128811

Mochida K, Saisho D, Hirayama T (2015) Crop improvement using life cycle datasets acquired under field conditions. Frontiers in Plant Sci 6:740. doi:10.3389/fpls.2015.00740.

Moretzsohn MC, Leoi L, Proite K, Guimaraes PM, Leal-Bertioli SCM, Gimenes MA, Martins WS, Valls JFM, Grattapaglia D, Bertioli DJ (2005) A microsatellite-based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae). Theor Appl Genet 111(6):1060–1071

Moretzsohn M, Barbosa A, Alves-Freitas D, Teixeira C, Leal-Bertioli S, Guimaraes P, Pereira R, Lopes C, Cavallari M, Valls J, Bertioli D, Gimenes M (2009) A linkage map for the B-genome of *Arachis* (Fabaceae) and its synteny to the A-genome. BMC Plant Biol 9(1):40

Morgante CV, Brasileiro ACM, Roberts PA, Guimaraes LA, Araujo ACG, Fonseca LN, Leal-Bertioli SCM, Bertioli DJ, Guimaraes PM (2013) A survey of genes involved in *Arachis*

*stenosperma* resistance to *Meloidogyne arenaria* race 1. Funct Plant Biol 40(12):1298–1309

Nagy E, Chu Y, Guo Y, Khanal S, Tang S, Li Y, Dong W, Timper P, Taylor C, Ozias-Akins P, Corley Holbrook C, Beilinson V, Nielsen N, Thomas Stalker H, Knapp S (2010) Recombination is suppressed in an alien introgression in peanut harboring Rma, a dominant root-knot nematode resistance gene. Mol Breeding 26(2):357–370

Nagy ED, Guo Y, Tang S, Bowers JE, Okashah RA, Taylor CA, Zhang D, Khanal S, Heesacker AF, Khalilian N, Farmer AD, Carrasquilla-Garcia N, Penmetsa RV, Cook D, Stalker HT, Nielsen N, Ozias-Akins P, Knapp SJ (2012) A high-density genetic map of *Arachis duranensis*, a diploid ancestor of cultivated peanut. BMC Genomics 13:469–469

Nautiyal P, Rajgopal K, Zala P, Pujari DS, Basu M, Dhadhal BA, Nandre BM (2008) Evaluation of wild *Arachis* species for abiotic stress tolerance: I. Thermal stress and leaf water relations. Euphytica 159(1–2):43–57

Nelson SC, Simpson CE, Starr JL (1989) Resistance to *Meloidogyne arenaria* in *Arachis* spp. germplasm. Supplement J Nematology 21:654–660

Ozias-Akins P, Gill R (2001) Progress in the development of tissue culture and transformation methods applicable to the production of transgenic peanut. Peanut Sci 28(2):123–131

Ozias-Akins P, Schnall JA, Anderson WF, Singsit C, Clemente TE, Adang MJ, Weissinger AK (1993) Regeneration of transgenic peanut plants from stably transformed embryogenic callus. Plant Sci 93(1–2):185–194

Pande S, Rao J (2001) Resistance of wild *Arachis* species to Late Leaf Spot and rust in greenhouse trials. Plant Dis 85:851–855

Pandey MK, Monyo E, Ozias-Akins P, Liang X, Guimarães P, Nigam SN, Upadhyaya HD, Janila P, Zhang X, Guo B, Cook DR, Bertioli DJ, Michelmore R, Varshney RK (2012) Advances in *Arachis* genomics for peanut improvement. Biotechnol Adv 30(3):639–651

Patel SS, Shah DB, Panchal HJ (2015) *De novo* transcriptome analysis of *Arachis hypogaea* L. (SRR1212866). OMICS Res 5

Peng Z, Gallo M, Tillman B, Rowland D, Wang J (2015) Molecular marker development from transcript sequences and germplasm evaluation for cultivated peanut (*Arachis hypogaea* L.). Mol Genet Genomics:1–19. doi:10.1007/s00438-015-1115-6

Porto MS, Pinheiro MPN, Batista VGL, dos Santos RC, de Albuquerque Melo Filho P, de Lima LM (2014) Plant promoters: an approach of structure and function. Mol Biotechnol 56(1):38–49

Proite K, Leal-Bertioli S, Bertioli D, Moretzsohn M, da Silva F, Martins N, Guimaraes P (2007) ESTs from a wild *Arachis* species for gene discovery and marker development. BMC Plant Biol 7(1):7

Proite K, Carneiro R, Falcao R, Gomes A, Leal-Bertioli S, Guimaraes P, Bertioli D (2008) Post-infection development and histopathology of *Meloidogyne arenaria* race 1 on *Arachis* spp. Plant Pathol 57(5):974–980. doi:10.1111/j.1365-3059.2008.01861.x

Rao S (2003) Nutrient balance and economics of integrated nutrient management in groundnut (*Arachis hypogaea* L.) - mustard (*Brassica juncea* L.). Madras Agric J 90(7–9):465–471

Reddy AS, Reddy LJ, Mallikarjuna N, Abdurahman MD, Reddy YV, Bramel PJ, Reddy DVR (2000) Identification of resistance to Peanut Bud Necrosis Virus (PBNV) in wild *Arachis* germplasm. Ann Appl Biol 137(2):135–139

Robertson D (2004) VIGS vectors for gene silencing: many targets, many tools. Annu Rev Plant Biol 55:495–519

Schmidt H, Gelhaus C, Latendorf T, Nebendahl M, Petersen A, Krause S, Leippe M, Becker W-M, Janssen O (2009) 2-D DIGE analysis of the proteome of extracts from peanut variants reveals striking differences in major allergen contents. Proteomics 9 (13):3507–3521

Seijo J, Lavia G, Fernandez A, Krapovickas A, Ducasse D, Moscone E (2004) Physical mapping of the 5S and 18S-25S rRNA genes by FISH as evidence that *Arachis duranensis* and *A. ipaensis* are the wild diploid progenitors of *A. hypogaea* (Leguminosae). Am J Bot 91:1294–1303

Senthil-Kumar M, Govind G, Kang L, Mysore KS, Udayakumar M (2007) Functional characterization of *Nicotiana benthamiana* homologs of peanut water deficit-induced genes by virus-induced gene silencing. Planta 225(3):523–539

Sharma S, Ansari M, Varaprasad K, Singh A, Reddy L (1999) Resistance to *Meloidogyne javanica* in wild *Arachis* species. Genet Resour Crop Evol 46(6):557–568

Sharma HC, Pampapathy G, Dhillon MK, Ridsdill-Smith JT (2005) Detached leaf assay to screen for host plant resistance to *Helicoverpa armigera*. J Econ Entomol 98(2):568–576

Sharma S, Upadhyaya HD, Varshney RK, Gowda C (2013) Pre-breeding for diversification of primary gene pool and genetic enhancement of grain legumes. Frontiers in Plant Sci 4. doi:10.3389/fpls.2013.00309

Shen Y, Zhiguo E, Zhang X, Liu Y, Chen Z (2015) Screening and transcriptome analysis of water deficiency tolerant germplasms in peanut (*Arachis hypogaea*). Acta Physiol Plant 37(5):1–9

Shirasawa K, Bertioli DJ, Varshney RK, Moretzsohn MC, Leal-Bertioli SCM, Thudi M, Pandey MK, Rami J-F, Foncéka D, Gowda MVC, Qin H, Guo B, Hong Y, Liang X, Hirakawa H, Tabata S, Isobe S (2013) Integrated consensus map of cultivated peanut and wild relatives reveals structures of the A and B genomes of *Arachis* and divergence of the legume genomes. DNA Res 20(2):173–184

Shukla P, Singh N, Kumar D, Vijayan S, Ahmed I, Kirti P (2014) Expression of a pathogen-induced cysteine protease (AdCP) in tapetum results in male sterility in transgenic tobacco. Funct Integr Genomics 14(2):307–317

Simpson CE, Starr JL (2001) Registration of 'COAN' peanut appreciation is expressed to the Texas Peanut Producers Board for their generous support of this research from 1988 to the present time. Registration by CSSA. Crop Sci 41(3):918

Singh A, Moss J (1984) Utilisation of wild relatives in the genetic improvement of *Arachis hypogaea* L. Theor Appl Genet 68(4):355–364

Singh NK, Kumar KRR, Kumar D, Shukla P, Kirti PB (2013) Characterization of a pathogen induced Thaumatin-Like protein gene *AdTLP* from *Arachis diogoi*, a wild peanut. PLoS ONE 8(12):e83963

Singsit C, Holbrook CC, Culbreath AK, Ozias-Akins P (1995) Progenies of an interspecific hybrid between *Arachis hypogaea* and *A.stenosperma*—pest resistance and molecular homogeneity. Euphytica 83(1):9–14

Sinharoy S, Saha S, Chaudhury SR, DasGupta M (2009) Transformed hairy roots of *Arachis hypogaea*: a tool for studying root nodule symbiosis in a non–infection thread legume of the Aeschynomeneae tribe. Mol Plant Microbe Interact 22(2):132–142

Sprent JI, Odee DW, Dakora FD (2009) African legumes: a vital but under-utilized resource. J Exp Bot:erp342

Stalker HT, Tallury SP, Ozias-Akins P, Bertioli D, Bertioli SCL (2013) The value of diploid peanut relatives for breeding and genomics. Peanut Sci 40(2):70–88

Subrahmanyam P, Moss JP, Rao VR (1983) Resistance to peanut rust in wild *Arachis* species. Plant Dis 67 (2):209–212

Sun Y, Wang Q, Li Z, Hou L, Dai S, Liu W (2013) Comparative proteomics of peanut gynophore development under dark and mechanical stimulation. J Proteome Res 12(12):5502–5511

Valls J, Simpson C (2005) New species of *Arachis* (Leguminosae) from Brazil, Paraguay and Bolivia. Bonplandia:35–63

Varshney RK, Roorkiwal M, Nguyen T (2013) Legume genomics: from genomic resources to molecular breeding. The Plant Genome 6(3):1–7

Varshney RK, Terauchi R, McCouch SR (2014) Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. PLoS Biol 12(6):e1001883

Varshney RK, Kudapa H, Pazhamala L, Chitikineni A, Thudi M, Bohra A, Gaur PM, Janila P, Fikre A, Kimurto P (2015) Translational genomics in agriculture: some examples in grain legumes. Crit Rev Plant Sci 34(1–3):169–194

Viquez OM, Konan KN, Dodo HW (2003) Structure and organization of the genomic clone of a major peanut allergen gene, Ara h 1. Mol Immunol 40(9):565–571

Wan X-R, Li L (2006) Regulation of ABA level and water-stress tolerance of *Arabidopsis* by ectopic expression of a peanut 9-cis-epoxycarotenoid dioxygenase gene. Biochem Bioph Res Co 347(4):1030–1038

Wang T, Zhang E, Chen X, Li L, Liang X (2010) Identification of seed proteins associated with resistance to pre-harvested aflatoxin contamination in peanut (*Arachis hypogaea* L). BMC Plant Biol 10(1):1

Wang Z, Yan S, Liu C, Chen F, Wang T (2012) Proteomic analysis reveals an aflatoxin-triggered immune response in cotyledons of *Arachis hypogaea* infected with *Aspergillus flavus*. J Proteome Res 11 (5):2739–2753

Wojciechowski MF (2003) Reconstructing the phylogeny of legumes (Leguminosae): an early 21st century perspective. Adv legume syst 10:5–35

Wu N, Matand K, Wu H, Li B, Li Y, Zhang X, He Z, Qian J, Liu X, Conley S, Bailey M, Acquaah G (2013) De novo next-generation sequencing, assembling and annotation of *Arachis hypogaea* L. Spanish botanical type whole plant transcriptome. Theor Appl Genet 126 (5):1145–1149

Xia H, Zhao C, Hou L, Li A, Zhao S, Bi Y, An J, Zhao Y, Wan S, Wang X (2013) Transcriptome profiling of peanut gynophores revealed global reprogramming of gene expression during early pod development in darkness. BMC Genom 14(1):517

Yin D, Wang Y, Zhang X, Li H, Lu X, Zhang J, Zhang W, Chen S (2013) De novo assembly of the peanut (*Arachis hypogaea* L.) seed transcriptome revealed candidate unigenes for oil accumulation pathways. PLoS ONE 8(9):e73767

Yu M, Liu F, Zhu W, Sun M, Liu J, Li X (2015) New features of triacylglycerol biosynthetic pathways of peanut seeds in early developmental stages. Funct Integr Genomics 15(6):707–716

Zhang J, Liang S, Duan J, Wang J, Chen S, Cheng Z, Zhang Q, Liang X, Li Y (2012) *De novo* assembly and characterisation of the transcriptome during seed development, and generation of genic-SSR markers in peanut (*Arachis hypogaea* L.). BMC Genom 13(1):90

Zhao C-Z, Xia H, Frazier T, Yao Y-Y, Bi Y-P, Li A-Q, Li M-J, Li C-S, Zhang B-H, Wang X-J (2010) Deep sequencing identifies novel and conserved microRNAs in peanuts (*Arachis hypogaea* L.). BMC Plant Biol 10(1):3

Zhao C, Zhao S, Hou L, Xia H, Wang J, Li C, Li A, Li T, Zhang X, Wang X (2015) Proteomics analysis reveals differentially activated pathways that operate in peanut gynophores at different developmental stages. BMC Plant Biol 15(1):188

Zhou Y, Yang P, Zhang F, Luo X, Xie J (2014) Analysis of promoter activity of the peanut (*Arachis hypogaea* L.) seed-specific allergen gene *Ara h 2.02* in transgenic *Arabidopsis*. Bothalia J 44(12):80–97

Zhu W, Zhang E, Li H, Chen X, Zhu F, Hong Y, Liao B, Liu S, Liang X (2013) Comparative proteomics analysis of developing peanut aerial and subterranean pods identifies pod swelling related proteins. J Proteomics 91:172–187

# Future Prospects for Peanut Improvement

# 11

Rajeev K. Varshney, Manish K. Pandey and Naveen Puppala

**Abstract**

Evolution in sequencing technologies led to reduction in costs and increase in speed for generating sequence data. The affordability of low-cost sequencing is expected to make other genotyping platforms obsolete in next couple of years. The concept of "single genome sequence" in a crop has evolved to sequencing of multiple genomes to assemble pangenomes. Sequencing combined with precise phenotyping of segregating populations and germplasm collections is expected to measure the accurate genetic diversity present in the germplasm as well as to identify the gene/nucleotide associated with the trait(s). It is time now to move toward using multi-parents populations from bi-parental populations for trait discovery and identify superior haplotypes. Availability of information on functional variation for genes controlling traits of interest will eventually help in manipulating genes more routinely using appropriate technologies such as marker-assisted selection/backcrossing, genomic selection, and genome editing. This chapter provides expected use of genome sequence and allied information on peanut for accelerating biology research as well as peanut improvement.

## 11.1 Introduction

The genus *Arachis,* containing 81 species arranged in nine taxonomic sections and variable genomes, face a huge genetic barrier for bringing exotic alleles to cultivated species from wild relatives (Pandey et al. 2012). Such genetic barrier not only limits the enhancing of genetic variation between different genepools, but also presents the challenge to researchers working on wild relatives for trait discovery and their

R.K. Varshney (✉) · M.K. Pandey
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India
e-mail: r.k.varshney@cgiar.org

R.K. Varshney
University of Western Australia (UWA), Crawley, Australia

N. Puppala
New Mexico State University, Agricultural Science Center, Clovis, USA



R.K. Varshney et al. (eds.), *The Peanut Genome*,
Compendium of Plant Genomes, DOI 10.1007/978-3-319-63935-2_11

deployment in breeding improved varieties. The existing trait dissection and breeding methodologies offer limited scope for large-scale modifications of genetic composition of lines. Therefore, advanced technologies must be developed and deployed to understand the complex and multiple genomes exist in genus *Arachis*. This will help in devising better and improved technologies and strategies for discovery of functional sequence variations in the genome associated with desirable traits in peanut. Availability of such information on desirable traits of peanut will facilitate faster alterations of multiple genes in the genome using modern molecular breeding technologies including genome editing.

The last decade has witnessed extraordinary progress in genome sequencing technologies leading to faster speed and huge reduction in cost per megabase. The high-throughput DNA sequencing technologies have achieved unprecedented scale of efficiency in sequencing followed by improved analytical tools for analyzing large-scale sequencing data to understand the complex biological problems. As a result, it has been possible to sequence several crop genomes including complex genomes (Goodwin et al. 2016). Genome sequence has become possible for AA and BB genome progenitor species of peanut (Bertioli et al. 2016; Chen et al. 2016). The onus, therefore, lies now on the peanut researchers for enhancing understanding of different *Arachis* genomes, identifying functional sequence variation followed by their deployment in peanut breeding using appropriate technologies and methods. Editors highlight some areas that can be addressed in coming years for accelerating peanut genetics and crop improvement.

## 11.2 Breaking Species/Section Barriers for Enhancing Genetic Base

*Arachis* genus is arranged in nine taxonomic sections carrying different genomes namely *Arachis* (A, B, D, F, G, and K), Trierectoides (E), Erectoides (E), Extranervosae (Ex), Triseminatae (T), Heteranthae (Am), Caulorrhizae (C), Procumbentes (E), and Rhizomatosae (R) (Krapovickas and Gregory 1994; Valls and Simpson 2005). Further, the cultivated peanut (*Arachis hypogaea* L.) can be divided into two subspecies (*hypogaea* and *fastigata*) based on the morpholocal differences in branching pattern and vegetative and reproductive axes. Based on inflorescence, pod, and seed characters, these subspecies can be further subclassified into six botanical varieties, i.e., *fastigata*, *vulgaris*, *peruviana*, and *aequatoriana* from subspecies *fastigata,* while *hypogaea* and *hirsuta* from subspecies *hypogaea.* In addition to above classifications, the peanut is also divided into different kinds of market types based on the seed size, plant growth type, and its applications in peanut industry, i.e., Spanish, Runner, Virginia, and Valencia. Currently, most of the breeding programs across the globe are engaged in developing improved varieties for one or two botanical and market types based on the demand in local and international markets. Genomics can play a major role in developing better understanding on different genomes of wild relatives, botanical, and market types so that improved varieties with specific features suitable to specific climatic conditions can be developed using genome-based breeding approaches.

## 11.3 Sequencing Reference Genome and Germplasm Collection for Developing Pangenomes and Hapmap

The cultivated peanut (*A. hypogaea*) is an allotetraploid ($2n = 4x = 40$) crop with two subgenomes. The Peanut Genome Consortium (PGC) with the collaboration of international partners developed first draft sequences of two progenitors of tetraploid cultivated peanut, representing A-genome (*Arachis duranensis*, accession V14167) and B-genome (*A. ipaensis,* accession K30076) (Bertioli et al. 2016). Another consortium namely Diploid Progenitor A-genome Sequencing Consortium (DPPAGSC) developed another draft sequence of A-genome

progenitor (*A. duranensis*, accession PI475845) (Chen et al. 2016). The effort by PGC is continued to develop a very high-quality tetraploid genome sequence for cultivated peanut by the end of this year.

Despite several efforts using cytogenetic and genetic studies, the level of genome diversity, genome evolution, and accurate phylogenetic relationship could not be established with a high level of precision and confidence in *Arachis* genus. Sequencing of mere two diploid progenitors and cultivated tetraploid genotypes do not represent the sequence variation present in the entire germplasm. It will be good to sequence all the genomes available in *Arachis* species including representative genotypes from each species from different sections. Although >15,000 accessions available in different genebanks at ICRISAT, USDA/ARS-Griffin, Georgia, USA, three important minicore collections namely ICRISAT MiniCore Collection (184 accessions), Chinese MiniCore Collection (298 accessions), and US MiniCore Collection (112 accessions) as well as three core collections namely ICRISAT Core Collection (1704 accessions), US Core Collection (831 accessions), and Chinese Core Collection (576 accessions) in addition to global Composite Collection (1000 accessions) can be started for resequencing in systematic manner. These efforts will provide core genome and pangenome helping in understanding genome evolution in a better way leading to answer key questions related to genome variations, evolution, phylogenetic relationship, and potential method of exchanging genome variations across *Arachis* genus. It will also be desirable to undertake phenotyping of those lines so that genome-wide association study (GWAS) at high-resolution level can be undertaken. Such an analysis will help to identify genes for traits of interest, superior haplotypes for a given gene (associated with traits of interest) and lines comprising suitable haplotypes for different genes. As the sequencing cost is reducing day by day, we are hopeful that global germplasm collections including in peanut will be sequenced in the long run.

## 11.4 Sequencing-Based Trait Dissection and Gene Discovery

In most of the current trait mapping studies, high-throughput genotyping using different kinds of SNP genotyping platforms is used for conducting linkage or GWAS studies. It has become possible now to use sequencing-based genotyping of the segregating populations. We anticipate that in the coming years, sequencing will be the approach of genotyping of mapping populations/germplasm sets for conducting high-resolution mapping and faster discovery of candidate genes for developing diagnostic markers for traits of interest. Most importantly, several analytical softwares have also become available for analyzing large datasets (Varshney et al. 2015). At present also, genotyping-by-sequencing (GBS), skim sequencing, and BSA-Seq/QTL-Seq approaches have been used for trait mapping (Pandey et al. 2016).

For high-resolution mapping, it is also essential to start using multi-parent genetic populations such as multi-parent advanced generation intercross (MAGIC), nested association mapping (NAM), and recombinant inbred advanced intercross line (RIAIL) populations (Morrell et al. 2012; Pandey et al. 2016). ICRISAT has developed three specialized MAGIC populations for aflatoxin resistance, drought tolerance, and nutritional and quality traits. Similarly, one NAM population each has been developed for Spanish and Virginia types by ICRISAT. These two types of multi-parent populations also provide opportunity to conduct joint linkage-association mapping (JLAM) in addition to linkage mapping. Such complex genetic populations often have several hundreds of individuals for genotyping and phenotyping similar to the majority of the association mapping panels. The low-cost sequencing will allow researchers in coming years to perform sequencing of complete mapping population/panel for conducting high-resolution trait mapping and candidate gene discovery. These developments will help in dissecting even the most complex traits such as drought tolerance, aflatoxin contamination, and allergens.

## 11.5 Next-Generation Breeding

The current molecular breeding approaches deploy genetic marker information using three major approaches namely marker-assisted backcrossing (MABC), marker-assistant recurrent selection (MARS), or genomic selection (GS) in crops including peanut (Varshney et al. 2013; Pandey et al. 2016). The successful example can be cited for improving selected traits such root-knot nematode (Chu et al. 2011), high oleic acid (Chu et al. 2011; Janila et al. 2016), and foliar fungal diseases (Varshney et al. 2014). However, there is no report available on deployment of other two approaches (MARS and GS) in peanut. Nevertheless, ICRISAT has taken some initiatives toward developing training population and completed its genotyping using 58 K SNP array developed recently (Pandey et al. 2017). Phenotyping of this training population is underway for different key agronomic traits to develop genomic selection model and initiation of GS breeding for some selected traits.

The other breeding approach is early generation screening (EGS) of large populations with at least markers for must-have traits. This will help enhancing selection intensity and in turn accelerate genetic gains in the breeding program. However, to deploy the EGS approach, it is essential to have diagnostic markers for majority/must-have traits.

As a result of large-scale genome resequencing projects, it will be possible to identify not just causal gene but also causal nucleotide for a given trait. In that scenario, it will be possible to undertake genome editing approach (Wood et al. 2011). In addition to adding the favorable alleles, genome editing also offers removal of deleterious alleles that have become available as a result of accelerated domestication of wild relatives. We envisage use of combination of EGS, GS, and genome editing in peanut in coming years. Of course, MABC should be continued to improve the elite/mega varieties for 1–2 traits for which varieties are deficient.

## 11.6 Conclusion

NGS technologies have undoubtedly accelerated the genomics research drastically leading to the generation of large data at reduced costs and less time. Sequencing of entire set of genomes, species, botanical varieties, and genbank germplasm will improve current understanding to devise novel strategies for harnessing the sequence diversity present across genepools. The low-cost sequencing will allow sequencing-based genotyping of large-scale populations/panels containing thousands of individuals for conducting high-resolution trait dissection and gene discovery, thereby making trait mapping more reliable and less time consuming. Given the speed of evolution in NGS technologies and advances in developing decision support tool, the next-generation breeding approaches will be used for peanut improvement.

## References

Bertioli DJ, Cannon SB, Froenicke L et al (2016) The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet 48:438–446

Chen X, Li H, Pandey MK et al (2016) Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens. PNAS 113:6785–6790

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

Goodwin S, PcPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17:333–351

Janila P, Pandey MK, Shasidhar Y et al (2016) Molecular breeding for introgression of fatty acid desaturase mutant alleles (ahFAD2A and ahFAD2B) enhances oil quality in high and low oil containing peanut genotypes. Plant Sci 242:203–213

Krapovickas A, Gregory WC (1994) Taxonomía del género *Arachis* (Leguminosae). Bonplandia 8:1–186

Morrell PL, Buckler ES, Ross-Ibarra J (2012) Crop genomics: advances and applications. Nat Rev Genet 13:85–96

Pandey MK, Monyo E, Ozias-Akins P, et al (2012) Advances in Arachis genomics for peanut improvement. Biotechnol Adv 30(3):639–651

Pandey MK, Agarwal G, Kale SM, et al (2017) Development and evaluation of a high density genotyping 'Axiom_*Arachis*' array with 58 K SNPs for accelerating genetics and breeding in groundnut. Sci Rep 7:40577

Pandey MK, Roorkiwal M, Singh VK et al (2016) Emerging genomic tools for legume breeding: current status and future prospects. Front Plant Sci 7:455

Valls JFM, Simpson CE (2005) New species of *Arachis* (Leguminosae) from Brazil, Paraguay and Bolivia. Bonplandia 14:35–63

Varshney RK, Mohan SM, Gaur PM et al (2013) Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. Biotechnol Adv 31:1120–1134

Varshney RK, Pandey MK, Janila P et al (2014) Marker assisted introgression of a QTL region to improve rust resistance in three elite and popular varieties of peanut (*Arachis hypogaea* L.). Theor Appl Genet 127:1771–1781

Varshney RK, Singh VK, Hickey J et al (2015) Analytical and decision support tools for genomics-assisted breeding. Trends Plant Sci. doi:10.1016/j.tplants.2015.10.018

Wood AJ, Lo T-W, Zeitler B, et al (2011) Targeted genome editing across species using ZFNs and TALENs. Science (New York, NY) 333(6040):307