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

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

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

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.

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

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 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