

Compendium of Plant Genomes  
*Series Editor: Chittaranjan Kole*

---

Rajeev K. Varshney  
Rachit K. Saxena  
Scott A. Jackson *Editors*

# The Pigeonpea Genome

---

# **Compendium of Plant Genomes**

## **Series editor**

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](mailto:ckole2012@gmail.com)

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

---

Rajeev K. Varshney · Rachit K. Saxena  
Scott A. Jackson  
Editors

# The Pigeonpea Genome

 Springer

*Editors*

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

Scott A. Jackson  
Center for Applied Genetic Technologies  
University of Georgia  
Athens, Georgia  
USA

Rachit K. Saxena  
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 978-3-319-63795-2          ISBN 978-3-319-63797-6 (eBook)  
DOI 10.1007/978-3-319-63797-6

Library of Congress Control Number: 2017951410

© Springer International Publishing AG 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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 F2 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 is 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 of interest not only 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 lifetime 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 to the Volume

The order of four nucleotides, i.e., adenine, cytosine, guanine, and thymine constitutes an organism's DNA or genome. Genome sequencing is to understand the order of nucleotides, or bases, in a genome which is responsible for basic behavior of every organism. With the current advances in the genome sequencing, the entire genomes have been decoded in a number of organisms. This has provided ways to manipulate genome constitutions for achieving desirable phenotypes in crops species as well. Such efforts in pigeonpea (*Cajanus cajan*) have been slow or negligible until onset of twenty-first century. The availability of reference genome sequence for cultivated pigeonpea has accelerated understanding the basic biology, understanding and deploying modern approaches for candidate gene discovery, and marker development for key traits.

Pigeonpea is an important legume crop in arid and semi-arid regions of the world. It provides nutritional food to the vegetarian families living in the various countries of the world. The early cultivated pigeonpea has been domesticated from its wild progenitor species, i.e., *C. cajanifolius* in central India around 3,000 years ago. This crop has unique feature of often cross-pollination behavior, which has been used in developing high yielding hybrids. Beginning of second decade of twenty-first century has been a starting phase of revolutionizing pigeonpea research as a number of high yielding hybrids have been released and reference genome sequence became available for candidate gene discovery, high-resolution trait mapping, and marker development and genomics-assisted breeding.

This book is well-timed in pigeonpea research as part of the genome compendium series for different crops. It contains 10 different chapters providing detailed overview on different aspects of botanical classification, genetics, genomics, and breeding of pigeonpea. This book not only provides information on recent advances on genome sequencing, genome architecture, genetic mapping, and marker identification but also presents future guidelines of research by deploying modern genomics tools in conjunction with breeding.

A total of 26 authors have contributed 10 chapters for this volume (see Appendix I). The editors of this volume are grateful to all the authors for their contribution in writing chapters and reviewers (see Appendix II) 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, Naresh Kumar and Jutta Lindenborn from Springer has been a great help in completion of this book and is 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, RKS is grateful to his wife (Shelly) and two young sons (Aniruddha and Madhav) for their help and moral support in doing editorial responsibilities in addition to research duties at ICRISAT. SJ also acknowledges his wife (Julie) for support.

Editors hope that their efforts in compiling the information on different aspects of pigeonpea will help the pigeonpea 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 pigeonpea research.

ICRISAT, India  
ICRISAT, India  
UGA, USA

Rajeev K. Varshney  
Rachit K. Saxena  
Scott A. Jackson

---

# Contents

<b>1</b>	<b>The Pigeonpea Genome: An Overview</b> . . . . .	1
	Rajeev K. Varshney, Rachit K. Saxena and Scott A. Jackson	
<b>2</b>	<b>Key Plant and Grain Characteristics and Their Importance in Breeding and Adaptation of Pigeonpea Cultivars</b> . . . . .	5
	K.B. Saxena, Rachit K. Saxena and Rajeev K. Varshney	
<b>3</b>	<b>Botanical Description of Pigeonpea</b> <b>[<i>Cajanus Cajan</i> (L.) Millsp.]</b> . . . . .	17
	C.V. Sameer Kumar, S.J. Satheesh Naik, Nidhi Mohan, Rachit K. Saxena and Rajeev K. Varshney	
<b>4</b>	<b>Wide Crossing Technology for Pigeonpea Improvement</b> . . . . .	31
	Nalini Mallikarjuna, Rachit K. Saxena, M. Byre Gowda and Rajeev K. Varshney	
<b>5</b>	<b>Modern Genomic Tools for Pigeonpea Improvement: Status and Prospects</b> . . . . .	41
	Abhishek Bohra, Shalini Pareek, Rintu Jha, Rachit K. Saxena, Indra P. Singh, Gaurav Pandey, Raj K. Mishra, Farindra Singh, Mayank Kaashyap, Rohit Joshi and Rajeev K. Varshney	
<b>6</b>	<b>Molecular Mapping of Genes and QTLs in Pigeonpea</b> . . . . .	55
	Reyazul Rouf Mir, Irshad Ahmad Rather, Mohd Ashraf Bhat, G.A. Parray and Rajeev K. Varshney	
<b>7</b>	<b>Germplasm Characterization and Trait Discovery</b> . . . . .	65
	Christopher P. Krieg, Mulualet T. Kassa and Eric J.B. von Wettberg	
<b>8</b>	<b>Whole-Genome Sequencing of Pigeonpea: Requirement, Background History, Current Status and Future Prospects for Crop Improvement</b> . . . . .	81
	Rachit K. Saxena and Rajeev K. Varshney	
<b>9</b>	<b>Sequencing Pigeonpea Genome</b> . . . . .	93
	Vikas K. Singh, Rachit K. Saxena and Rajeev K. Varshney	
<b>10</b>	<b>Future Prospects</b> . . . . .	99
	Rajeev K. Varshney, Rachit K. Saxena and Scott A. Jackson	