

Compendium of Plant Genomes
Series Editor: Chittaranjan Kole

Malali Gowda
H. B. Mahesh
Chittaranjan Kole *Editors*

The Sandalwood Genome

Compendium of Plant Genomes

Series Editor

Chittaranjan Kole, President, International Climate Resilient Crop Genomics Consortium (ICRCGC), President, International Phytomedomics & Nutriomics Consortium (IPNC) and President, Genome India International (GII), Kolkata, 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 100 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.

More information about this series at <https://link.springer.com/bookseries/11805>

Malali Gowda • H. B. Mahesh •
Chittaranjan Kole
Editors

The Sandalwood Genome

 Springer

Editors

Malali Gowda
Bengaluru Genomics Center
Bangalore, India

H. B. Mahesh
Genetics and Plant Breeding
University of Agricultural Sciences
Bangalore, India

Chittaranjan Kole
International Climate Resilient Crop
Genomics Consortium (ICRCGC)
International Phytomedomics
and Nutriomics Consortium (IPNC)
Genome India International (GII)
Kolkata, India

ICAR-National Institute for Plant
Biotechnology
Pusa, New Delhi, India

ISSN 2199-4781

ISSN 2199-479X (electronic)

Compendium of Plant Genomes

ISBN 978-3-030-93393-7

ISBN 978-3-030-93394-4 (eBook)

<https://doi.org/10.1007/978-3-030-93394-4>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer
Nature Switzerland AG 2022

This work is subject to copyright. All rights are solely and exclusively licensed 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, expressed 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.

This Springer imprint is published by the registered company Springer Nature Switzerland 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 the 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 the 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 the 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, the 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. The 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, the 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 are 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 to both 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, the 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, particularly Dr. Christina Eckey and Dr. Jutta Lindenborn for the earlier set of volumes and presently Ing. Zuzana Bernhart for all their timely help and support.

I always had to set aside additional hours to edit books beside 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.

New Delhi, India

Chittaranjan Kole

Preface

Indian sandalwood is one of the most valuable trees in the world. The essential oil extracted from heartwood has gained a significant place in perfumery industries. Unfortunately, anthropogenic activity has caused a gradual decrease in the production of sandalwood in India. Recent data on production of sandalwood in India showed a declining trend and it is considered as endangered due to over-harvesting of roots of this tree. There is a limited research done to understand the genetic diversity of sandalwood using genome-wide markers. In addition, lack of genomic resources for this highly prized tree, was an impediment for the diversity studies and its conservation in the natural habitat. The genomic studies in sandalwood have primarily focused on understanding the molecular mechanism governing santalol oil content. However, understanding the sandalwood at whole-genome level was unattended till recent past. Our motivation is to compile the genomic, transcriptomic, and proteomic information into a book. This book can serve as a ready reckoner for sandalwood researchers. Recent advances in genome sequencing and transcriptome profiling revealed a greater insight on the transformation of sandal with gene(s) resistant to various sandal diseases and for metabolic engineering to increase or modify the essential oil yield. The characterization and functional annotation of genes responsible for the production of essential oil in sandalwood, is one of the major areas of focus due to its economic importance in preparation of various perfumes and cosmetics.

Whole genome sequence of different sandalwood ecotypes, its comparative studies, and multiomic information provided in this book holds wide implication for future research and has potential to reach out to the global scientific community. Very importantly, the combination of information from multi-omics resources will enable trait-specific breeding and conservation of genetic resources of *S. album*. The molecular understanding of the sesquiterpene pathway will have industrial implications through biomarker discovery and metabolic engineering in microbial system for santalol production.

Bangalore, India
Bangalore, India
Kolkata, India

H. B. Mahesh
Malali Gowda
Chittaranjan Kole

Contents

1 Sandalwood: The King of Aroma in Plant Kingdom	1
H. B. Mahesh and Malali Gowda	
2 Origin, Tradition and Cultural Importance of Sandalwood in India and Its Role in Conservation of Natural Populations	7
C. Sandeep and T. N. Manohara	
3 Sandalwood: Traditional Knowledge from Ayurveda	19
S. N. Venugopalan Nair, Tabassum Ishrath Fathima, and Shilpa Naveen	
4 Genomic Resources in Indian Sandalwood: Implications in Conservation and Breeding	31
Eswaran Muthulakshmi, Chandramouli K. Madhuvanathi, and Modhumita Ghosh Dasgupta	
5 Whole Genome Sequence of Sandalwood and Its Comparative Study	47
H. B. Mahesh, Meghana Deepak Shirke, K. G. Manasa, and Malali Gowda	
6 Mining of SSRs and SNPs in Sandalwood Genome	57
H. B. Mahesh, K. G. Manasa, Meghana Deepak Shirke, and N. R. Raghavendra	
7 Sandalwood Fragrance Pathway and Its Engineering for Sustainable Production of High-Value Santalols	65
Ananth K. Narayanan, Sudharshan SJ, and Dinesh A. Nagegowda	
8 Sandalwood Essential Oil and Its Phyto-Chemistry	83
N. R. Raghavendra and H. B. Mahesh	
9 Scope of Proteogenomic Analysis for Indian Sandalwood (<i>Santalum Album</i>)	89
Pratigya Subba and Thottethodi Subrahmanya Keshava Prasad	
10 Future Prospects of Sandalwood Research	99
H. B. Mahesh, K. G. Manasa, M. K. Prasannakumar, Meghana Deepak Shirke, Chittaranjan Kole, and Malali Gowda	

Contributors

Tabassum Iskrath Fathima Centre for Traditional Knowledge & Informatics, The University of Trans-Disciplinary Health Sciences and Technology (TDU), Bengaluru, India

Modhumita Ghosh Dasgupta Institute of Forest Genetics and Tree Breeding, R.S. Puram, Coimbatore, India

Malali Gowda The University of Trans-Disciplinary Health Sciences and Technology (TDU), Jarakabande Kaval, Bengaluru, India;
Bengaluru Genomics Center, Yelahanka, Bengaluru, India

Chittaranjan Kole ICAR-National Research Center on Plant Biotechnology, Pusa, New Delhi, India

Chandramouli K. Madhuvanathi Institute of Forest Genetics and Tree Breeding, R.S. Puram, Coimbatore, India

H. B. Mahesh Department of Genetics and Plant Breeding, College of Agriculture, University of Agricultural Sciences, Mandya, Bengaluru, India

K. G. Manasa Department of Genetics and Plant Breeding, College of Agriculture, Mandya, University of Agricultural Sciences, Bangalore, India

T. N. Manohara Silviculture and Forest Management Division, Institute of Wood Science and Technology, Malleswaram, Bengaluru, India

Eswaran Muthulakshmi Institute of Forest Genetics and Tree Breeding, R.S. Puram, Coimbatore, India

Dinesh A. Nagegowda Molecular Plant Biology and Biotechnology Lab, CSIR-Central Institute of Medicinal and Aromatic Plants, Bengaluru, India; Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India

S. N. Venugopalan Nair Centre for Traditional Knowledge & Informatics, The University of Trans-Disciplinary Health Sciences and Technology (TDU), Bengaluru, India

Ananth K. Narayanan Molecular Plant Biology and Biotechnology Lab, CSIR-Central Institute of Medicinal and Aromatic Plants, Bengaluru, India; Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India