Compendium of Plant Genomes
Series Editor: Chittaranjan Kole

Minami Matsui
Keng-See Chow *Editors*

The Rubber Tree Genome

Springer

Compendium of Plant Genomes

Series Editor

Chittaranjan Kole, Raja Ramanna Fellow, Government of India, ICAR-National Research Center on Plant Biotechnology, Pusa, New Delhi, 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.

Interested in editing a volume on a crop or model plant? Please contact Prof. C. Kole, Series Editor, at ckoleorg@gmail.com

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

Minami Matsui • Keng-See Chow
Editors

The Rubber Tree Genome

Springer

Editors
Minami Matsui
RIKEN Center for Sustainable Resource Science
Yokohama, Kanagawa, Japan

Keng-See Chow
Academy of Sciences Malaysia
Subang Jaya, Selangor, Malaysia

ISSN 2199-4781 ISSN 2199-479X (electronic)
Compendium of Plant Genomes
ISBN 978-3-030-42257-8 ISBN 978-3-030-42258-5 (eBook)
https://doi.org/10.1007/978-3-030-42258-5

© Springer Nature Switzerland AG 2020
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, 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 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 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.

Kalyani, India Chittaranjan Kole

Preface

Hevea brasiliensis Müll. Arg., also known as the para rubber tree, is the sole species that is cultivated for the commercial production of natural rubber. Natural rubber (or *cis*-polyisoprene) is synthesized in the latex of this tree and harvested by systematic tapping of the tree trunk. The productive life of the rubber tree may extend beyond 20 years from first bark tapping in plantations managed under optimal agronomic practices. The genus *Hevea* (2n = 36) comprises 10 species, however only the *brasiliensis* species has been commercialized on a wide scale, primarily in Southeast Asian countries. The success of rubber production facilitated the growth of the automobile, medical, aviation, electronics, and building and construction industries, not to mention the manufacturing of a substantial range of rubber products essential to daily human life. Southeast Asian countries generate more than 95% of the global natural rubber production to meet increasing world demand. Through the effort of breeding and selection, the production capacity of modern rubber clones has reached more than four times that of the original parental trees. Nonetheless, the plantation industry is still facing the threat of diseases and the limitations of a narrow gene pool for breeding.

The significance of the rubber tree as a world crop warrants investment into genome-based technologies to deepen the current understanding of tree physiologies, particularly in latex, which affect production capacity. The rubber tree has accumulated a wealth of fundamental biology based on conventional scientific approaches spanning most of the twentieth century. Genomics and biotechnology approaches complement the existing knowledgebase and have great potential in discovering genes, proteins, and other regulatory components that could be engineered to enhance tree productivity. The period 2013–2017 was pivotal for rubber genome sequencing: the first draft genome sequence was published in 2013, followed by the release of three higher quality genome sequences of rubber tree clones using advanced sequencing technologies. Concurrently, genome-based investigations flourished as evidenced by the large number of publications by members of the rubber tree research community, particularly among the molecular biologists, physiologists, geneticists, and breeders.

In this book, we have compiled recent genome-related research, beginning with the experiences of the first three research groups in sequencing the genomes of rubber clones Reyan 7-33-97, RRIM 600, and BPM 24. The importance of harnessing benefits from genome information is reflected in chapters discussing the development of markers, genetic linkage maps, marker-trait association, and databases to facilitate data sharing and utilization. Applications of transcriptome and proteome analyses of major metabolic pathways involved in rubber biosynthesis and latex yield are also included. This book has drawn on the experience of contributors with a broad range of expertise in rubber tree research and/or background in genomic technologies. Nonetheless, we also recognize the value of parallel developments in other plants which have similarly accumulated sequence resources, namely, the Euphorbiaceae species such as *Manihot esculenta* (cassava), *Ricinus communis* (castor bean) and *Jatropha curcas* (physic nut), and alternative rubber-producing species such as *Taraxacum kok-saghyz*. We hope that this book will serve to record genome sequencing as a milestone in rubber tree research, and also the scientific investigations made during the initial phase of the research community's engagement with genome data.

Genomics is rapidly evolving and so will the directions of rubber tree genomics as more and more cutting-edge technologies become available. In the continuing quest for high-resolution genomes, we can expect to see publications of chromosome-level genome assemblies, high-density genetic maps, and tools for genome selection in the near future. Given the potential of inter- and intra-species genetic diversity for molecular breeding, we may also expect additional genomes of different *Hevea* species and cultivated genotypes to be sequenced. Insights from comparative genomics are likely to promote closer cooperation between genome scientists in the fields of *Hevea* and non-*Hevea* rubber research.

The rubber tree research community is relatively small compared to many major agricultural crops. The International Rubber Research and Development Board (IRRDB) has played an important role in promoting links between genomics and biotechnology researchers from different natural rubber-related organizations. In the course of preparing this book, we have benefited from some of the annual activities and wish to thank the IRRDB for facilitating such useful meetings and discussions. Finally, we also thank Prof. Chittaranjan Kole and the Springer team for their valuable advice and kind assistance.

Members from three rubber genome groups and the IRRDB Molecular Biology and Physiology Specialist Group gathered at the Sungai Buloh Research Station, Malaysian Rubber Board, Selangor on April 9, 2019.
From left: Azlina Bahari (MRB), Yuko Makita (RIKEN), Han Cheng (CATAS), Minami Matsui (RIKEN), Roslinda Sajari (MRB), Thitaporn Phumichai (RAOT), Keng-See Chow (MRB), Sithichoke Tangphatsornruang (BIOTEC).
MRB: Malaysian Rubber Board, Malaysia.
RIKEN: RIkagaku KENkyusho, Japan.
CATAS: Chinese Academy of Tropical Agricultural Sciences, China.
RAOT: Rubber Authority of Thailand, Thailand.
BIOTEC: National Center for Genetic Engineering and Biotechnology, Thailand.

Yokohama, Japan — Minami Matsui, D.Sci
Subang Jaya, Malaysia — Keng-See Chow, Ph.D.
January 2020

Contents

Contributors

Azlina Bahari Genetic Resources and Improvement Unit, Production Development Division, Malaysian Rubber Board, Sungai Buloh, Selangor, Malaysia

Han Cheng Key Laboratory of Rubber Biology, Ministry of Agriculture, Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan, People's Republic of China

Keng-See Chow Academy of Sciences Malaysia, Selangor, Malaysia

Katrina Cornish Departments of Horticulture and Crop Science, and Food, Agricultural and Biological Engineering, Ohio Agricultural Research and Development Center, College of Food, Agricultural and Environmental Science, The Ohio State University, Wooster, OH, USA

Longjun Dai Key Laboratory of Biology and Genetic Resources of Rubber Tree, Ministry of Agriculture and Rural Affairs, Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou, People's Republic of China

Muhammad Akbar Abdul Ghaffar Latex Harvesting Technologies and Physiology Unit, Production Development Division, Malaysian Rubber Board, Sungai Buloh, Selangor, Malaysia

Huasun Huang Key Laboratory of Rubber Biology, Ministry of Agriculture, Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan, People's Republic of China

Mika Kawashima Synthetic Genomics Research Group, RIKEN Center for Sustainable Resource Science (CSRS), Yokohama, Kanagawa, Japan

Yukio Kurihara Synthetic Genomics Research Group, RIKEN Center for Sustainable Resource Science (CSRS), Yokohama, Kanagawa, Japan

Nyok-Sean Lau Centre for Chemical Biology, Universiti Sains Malaysia, Bayan Lepas, Penang, Malaysia;
Synthetic Genomics Research Group, RIKEN Center for Sustainable Resource Science (CSRS), Yokohama, Kanagawa, Japan

Dejun Li Key Laboratory of Biology and Genetic Resources of Rubber Tree, Ministry of Agriculture and Rural Affairs, Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou, People's Republic of China

Yuko Makita Synthetic Genomics Research Group, RIKEN Center for Sustainable Resource Science (CSRS), Yokohama, Kanagawa, Japan

David F. Marshall Scotland's Rural College (SRUC), Edinburgh, Scotland, UK

Minami Matsui Synthetic Genomics Research Group, RIKEN Center for Sustainable Resource Science (CSRS), Yokohama, Kanagawa, Japan

Ahmad Sofiman Othman Centre for Chemical Biology, Universiti Sains Malaysia, Bayan Lepas, Penang, Malaysia;
School of Biological Sciences, Universiti Sains Malaysia, Minden, Penang, Malaysia

Wirulda Pootakham National Omics Center, National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Science Park, Thailand, Khlong Nueng, Pathumthani, Thailand

Jeremy R. Shearman National Omics Center, National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Science Park, Thailand, Khlong Nueng, Pathumthani, Thailand

Chaorong Tang School of Agricultural Science, Hainan University, Haikou, Hainan, People's Republic of China

Sithichoke Tangphatsornruang National Omics Center, National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Science Park, Thailand, Khlong Nueng, Pathumthani, Thailand

Mark A. Taylor The James Hutton Institute, Invergowrie, Dundee, Scotland, UK

Shaohua Wu Key Laboratory of Biology and Genetic Resources of Rubber Tree, Ministry of Agriculture and Rural Affairs, Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou, People's Republic of China

Hoong-Yeet Yeang Academy of Sciences Malaysia, Selangor, Malaysia

Abbreviations

4DTv	Transversion rate at fourfold degenerate sites
ABA	Abscisic acid
ACAT	Acetyl-CoA acetyltransferase
ACO	ACC oxidase
ACS	1-aminocyclopropane-1-carboxylic acid synthase
AFLP	Amplified fragment length polymorphism
AFSM	Amplified fragment SNP and methylation
ATAC-seq	Assay for transposase-accessible chromatin using sequencing
atm	Atmosphere
BAC	Bacterial artificial chromosome
BLASTP	Basic local alignment search tool (protein)
bp	Base pairs
bZIP	Basic leucine zipper transcription family
CAGE	Cap analysis gene expression
CDS	Coding sequence
ChIP-seq	Chromatin immunoprecipitation sequencing
cM	Centimorgan
COP	Coat protein complex
COR	Coronatine
CPT	*cis*-prenyltransferase
CPTL/CPTBP	CPT-like/CPT-binding protein
CTC	CAGE tag cluster
DC	Donor clone
DEG	Differentially expressed gene
DEP	Differentially expressed protein
DNase-seq	DNase I hypersensitive sites sequencing
dpi	Days post-infection
DXR	1-deoxy-D-xylulose 5-phosphate reductoisomerase
DXS	1-deoxy-D-xylulose 5-phosphate synthase
EIN2	Ethylene insensitive 2
ER	Endoplasmic reticulum
EST	Expressed sequence tag
ET	Ethylene
ETR	Ethylene resistance 1

EU	Extension unit
FDE	Filtered differentially expressed
FL-cDNA	Full-length cDNA
FPKM	Fragments per kilobase of transcript per million fragments mapped
FPP	Farnesyl pyrophosphate
FPPS	Farnesyl diphosphate synthase
FPS	Farnesyl diphosphate synthase
Gb	Giga base pairs
GBS	Genotyping-by-sequencing
GGPS	Geranylgeranyl pyrophosphate synthase
GO	Gene Ontology
GPPS	Geranyl diphosphate synthase
GPS	Geranyl pyrophosphate synthase
GS	Genome selection
GWAS	Genome-wide association study
HiC	High-throughput chromosome conformation capture
HMGR	3-hydroxy-3-methylglutaryl-CoA reductase
HMGS	3-hydroxy-3-methylglutaryl-CoA synthase
HMM	Hidden Markov model
hpi	Hours post-infection
HRBP	HRT1-REF bridging protein
IDP	Isopentenyl diphosphate
Indel	Insertion and deletion
Inr	Initiator element
IPP	Isopentenyl pyrophosphate
IPPI	Isopentenyl diphosphate isomerase
IRRDB	International Rubber Research and Development Board
Iso-Seq	Isoform sequencing
JA	Jasmonate
JAZ	Jasmonate ZIM-domain
JC	Juvenile clones
Kb	Kilobase pairs
KEGG	Kyoto Encyclopedia of Genes and Genomes
kg/ha/yr	Kilograms/hectare/year
$K_m^{IPP\text{-}Mg}$	Binding constant for the IPP-magnesium substrate
LA	Linolenic acid
LC-MS/MS	Liquid chromatography with tandem mass spectrometry
LG	Linkage group
LRP	Large rubber particle
LTR	Long-terminal repeat
LTR-RT	Long-terminal repeat retrotransposon
MALDI-TOF	Matrix-assisted laser desorption/ionization-time of flight
MAPK	Mitogen-activated protein kinase
MAS	Marker-assisted selection
Mb	Megabase pairs
MEP	2-C-methyl-D-erythritol 4-phosphate

miRNA	Micro ribonucleic acid
MNase-seq	Micrococcal nuclease-sequencing
mRNA	Messenger ribonucleic acid
MVA	Mevalonate
MWD	Molecular weight distribution
mya	Million years ago
NCBI	The National Center for Biotechnology Information
ncRNA	Non-coding RNA
NgBR	Nogo-B receptor
NGS	Next-generation sequencing
NMD	Nonsense-mediated decay
Nr	Non-redundant
NRS	Non-redundant sequence
ORF	Open reading frame
PAML	Phylogenetic analysis by maximum likelihood
PCD	Programmed cell death
PDC	Pyruvate dehydrogenase complex
PE	Paired-end
QTL	Quantitative trait locus
RACE	Rapid amplification of cDNA ends
RAPD	Random amplified polymorphic DNA
RBIP	Rubber biosynthesis inhibitor protein
RBSP	Rubber biosynthesis stimulator protein
REF	Rubber elongation factor
RFLP	Restriction fragment length polymorphism
RNA-seq	RNA sequencing
ROS	Reactive oxygen species
RPKM	Reads per kilobase of transcript per million mapped reads
rRNA	Ribosomal RNA
RT	Rubber transferase
RT-ase	Rubber transferase
SALB	South American Leaf Blight
SAMS	S-adenosyl-L-methionine synthase
SDR	Short-chain dehydrogenase/reductase
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	Scanning electron microscope
SMRT	Single-molecule real time
snoRNA	Small nucleolar ribonucleic acid
SNP	Single nucleotide polymorphism
snRNA	Small nuclear RNA
SOD	Superoxide dismutase
SRA	Sequence read archive
SRPP	Small rubber particle protein
SRPRNA	Signal recognition particle RNA
SSH	Suppression subtractive hybridization
SSR	Simple sequence repeat

TE	Transposable element
TEM	Transmission electron microscopy
TF	Transcription factor
TFBS	Transcription factor-binding site
TGN	Trans-Golgi network
tmRNA	Transfer-messenger RNA
TPD	Tapping panel dryness
TPM	Tags per million
tRNA	Transfer RNA
TSA	Trichostatin A
TSS	Transcription start site
uORF	Upstream ORF
UPP	Ubiquitin proteasome pathway
WGCNA	Weighted gene co-expression network analysis
WGS	Whole-genome shotgun
WRP	Detergent-washed rubber particle