Compendium of Plant Genomes
Series Editor: Chittaranjan Kole

Jiang Xu
Tae-Jin Yang
Hao-yu Hu *Editors*

The Ginseng 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

Jiang Xu • Tae-Jin Yang •
Hao-yu Hu
Editors

The Ginseng Genome

Springer

Editors
Jiang Xu
Institution of Chinese Materia Medica
China Academy of Chinese Medical Sciences
Beijing, China

Hao-yu Hu
Institution of Chinese Materia Medica
China Academy of Chinese Medical Sciences
Beijing, China

Tae-Jin Yang
Department of Plant Science, CALS
Seoul National University
Seoul, Korea (Republic of)

ISSN 2199-4781 ISSN 2199-479X (electronic)
Compendium of Plant Genomes
ISBN 978-3-030-30346-4 ISBN 978-3-030-30347-1 (eBook)
https://doi.org/10.1007/978-3-030-30347-1

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

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.

New Delhi, India Chittaranjan Kole

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

Chittaranjan Kole

Preface

Panax ginseng C. A. Mey, a deciduous perennial plant belonging to the Araliaceae family, has been clinically used as a precious herbal medicine for several millennia in East Asia. The name ginseng was translated from the pronunciation of the Chinese words "Ren shen." Modern pharmacological research confirmed that ginsenosides, the major bioactive compound of *P. ginseng*, exhibit multiple therapeutic activities. These activities include antitumor, antihypertensive, antivirus, and immune modulatory activities [3]. Therefore, *P. ginseng* is used as a general tonic or adatogen to promote longevity, particularly in China, Korea, and Japan. Currently, two genomes as well as several transcriptome data have been released for this valuable crop. Here, we present this book for readers to review the current achievements of omics of ginseng.

This book comprised with 14 chapters can be generally separated into three parts.

The first parts briefs about the background of ginseng studies. In the first two chapters, the authors introduced *Panax ginseng* from its origin, distribution, germplasm, cultivation, and medicinal properties. As molecular identification is the most early molecular biological tools used in ginseng industry, we introduced some work in this field in Chap. 3. In Chap. 4, the author introduced the progress of ginseng breeding which includes the breeding methods and mainstream varieties characteristics.

The second part focused on the genome-wide research of ginseng. Cytogenetics once provided the image base for ginseng genetic research. Chapter 5 not only reviewed the historical achievements of *P. ginseng* cytogenetics but also discussed its future perspectives in the post-genomics era. In Chaps. 6 and 7, the authors introduced ginseng genome from two aspects—the metabolic pathway and the evolution. These two chapters will give the readers a panorama view of the genome information of this valuable medicinal plant. And besides the nuclear genome, cytoplasm genome is an unignoring part of plant genome. Chapter 8 presented chloroplast genome variation among *Panax* species and individuals. The authors also introduced the application of chloroplast genome in ginseng authentication. In the ninth chapter, the book focused on the transcriptome sequences and their technological advancements for the *Panax* species. The history of ginseng transcriptome research is far more than genome itself. So in this chapter, more technologies will be presented to our readers, including ESTs, 454 and up-to-date long-read single-molecule real-time technique. Chapter 10 discussed the metabolite pathways and metabolic dynamics. This chapter will help readers to joint to the metabolics and genes together. Chapter 11 gave a comprehensive overview of ginseng genome database, tools, and its broad utilities. Tools provided in this chapter will help readers use current achievements of ginseng omics even without profound informatics skills. *Panax notoginseng* is another important medicinal plant in genus *Panax*; here, the authors also introduced its genome characters in Chap. 12.

In the third part, we introduced the progress of ginsenosides biosynthesis and the microbiome for ginseng medicine. Chapter 13 traced the recent progress of synthetic biology for ginsenoside production, including the elucidation of biosynthetic pathways and construction of cell factories for both natural and non-natural ginsenosides. These works will promote the industrialization of ginsenosides. In the last chapter, the authors summarized the role of gut microbiota in mediating the metabolism and enhanced bioavailability of ginseng.

This book covered the mainly achievements of ginseng research in omic times. We believe the information contained in this book will benefit the researchers, students, and producers of ginseng. We also sincerely appreciate the help of our editorial group from Springer Nature.

Beijing, China — Jiang Xu
Seoul, Korea (Republic of) — Tae-Jin Yang
Beijing, China — Hao-yu Hu

Contents

Contributors

Jae Youl Cho Department of Integrative Biotechnology, Sungkyunkwan University, Suwon, Republic of Korea

Vo Ngoc Linh Giang Department of Plant Science, Plant Genomics and Breeding Institute, Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul, Korea

Zhang Guang-hui National-Local, Joint Engineering Research Center on Germplasm Utilization and Innovation of Chinese Medicinal Materials in Southwest China, Yunnan Agricultural University, Kunming, China

Jianping Han Engineering Research Center of Tradition Chinese Medicine Resource, Ministry of Education, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Hu Hao-yu Key Laboratory of Beijing for Identification and Safety Evaluation of Chinese Medicine, Institution of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China

Dong-Yun Hyun Division of Ginseng, Department of Herbal Crop Research, National Institute of Horticultural and Herbal Science, Rural Development Administration, Eumseong, Korea

Woojong Jang Department of Plant Science, Plant Genomics and Breeding Institute, Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul, Korea

Murukarthick Jayakodi Department of Agriculture, Forestry and Bioresources, Plant Genomics and Breeding Institute, College of Agriculture and Life Sciences, Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul, Korea;
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Seeland, Germany

Xu Jiang Key Laboratory of Beijing for Identification and Safety Evaluation of Chinese Medicine, Institution of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China

Ki-Hong Jung Graduate School of Biotechnology and Crop Biotech Institute, Kyung Hee University, Yongin, Korea

Kondou Kenji TSUMURA & CO, Tokyo, Japan

Hyun Hee Kim Department of Life Science, Chromosome Research Institute, Sahmyook University, Seoul, Republic of Korea

Jang-Uk Kim Division of Ginseng, Department of Herbal Crop Research, National Institute of Horticultural and Herbal Science, Rural Development Administration, Eumseong, Korea

Nam-Hoon Kim Department of Agriculture, Forestry and Bioresources, Plant Genomics and Breeding Institute, College of Agriculture and Life Sciences, Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul, Korea;
PHYZEN Genomics Institute, Seongnam-si, Gyeonggi-do, Korea

Young-Chang Kim Division of Ginseng, Department of Herbal Crop Research, National Institute of Horticultural and Herbal Science, Rural Development Administration, Eumseong, Korea

Yu-Jin Kim Department of Life Science and Environmental Biochemistry, Pusan National University, Miryang, Republic of Korea

Chaojing Li CAS-Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China;
University of Chinese Academy of Sciences, Beijing, China

Xiangyan Li Key Laboratory of Active Substances and Biological Mechanisms of Ginseng Efficacy, Ministry of Education, Jilin Provincial Key Laboratory of BioMacromolecules of Chinese Medicine, Jilin Ginseng Academy, Changchun University of Chinese Medicine, Changchun, Jilin, China

Yang Liu Engineering Research Center of Tradition Chinese Medicine Resource, Ministry of Education, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Zhao Lizi The Engineering Research Institute of Agriculture and Forestry, Ludong University, Yantai, Shandong, P.R. China

Ivan Mijakovic Systems and Synthetic Biology, Chalmers University of Technology, Göteborg, Sweden;
The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark

Padmanaban Mohanan Plant Genomics and Breeding Institute, Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul, Korea

Hyun-Seung Park Department of Plant Science, Plant Genomics and Breeding Institute, Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul, Korea

Shadi Rahimi Systems and Synthetic Biology, Chalmers University of Technology, Göteborg, Sweden

Zubair Ahmed Ratan Department of Biomedical Engineering, Khulna University of Engineering and Technology, Khulna, Bangladesh

Yang Sheng-chao National-Local, Joint Engineering Research Center on Germplasm Utilization and Innovation of Chinese Medicinal Materials in Southwest China, Yunnan Agricultural University, Kunming, China

Hyeonah Shim Department of Agriculture, Forestry and Bioresources, Plant Genomics and Breeding Institute, Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul, Korea

Xiao Shuiming Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China

Nomar Espinosa Waminal Department of Life Science, Chromosome Research Institute, Sahmyook University, Seoul, Republic of Korea

Gang Wang Engineering Research Center of Tradition Chinese Medicine Resource, Ministry of Education, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Pingping Wang CAS-Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Zhang Xiaoyan Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China

Xing Yan CAS-Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Chengshuai Yang CAS-Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Deok-Chun Yang Graduate School of Biotechnology and Department of Oriental Medicinal Biotechnology, College of Life Sciences, Kyung Hee University, Yongin, Republic of Korea

Tae-Jin Yang Department of Agriculture, Forestry and Bioresources, Plant Genomics and Breeding Institute, Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul, Korea

Wenzhi Yang Tianjin State Key Laboratory of Modern Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin, China

Xu Yonghua Jilin Agricultural University, Changchun City, Jilin Province, P.R. China

Lu Yu CAS-Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Dabing Zhang Joint International Research Laboratory of Metabolic and Developmental Sciences, Shanghai Jiao Tong University–University of Adelaide Joint Centre for Agriculture and Health, State Key Laboratory of Hybrid Rice, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China;
Graduate School of Biotechnology and Crop Biotech Institute, Kyung Hee University, Yongin, Korea

Naiwu Zhang China Medico Coperation, Tianjin, China

Daqing Zhao Key Laboratory of Active Substances and Biological Mechanisms of Ginseng Efficacy, Ministry of Education, Jilin Provincial Key Laboratory of BioMacromolecules of Chinese Medicine, Jilin Ginseng Academy, Changchun University of Chinese Medicine, Changchun, Jilin, China

Xing Zhi-han Key Laboratory of Beijing for Identification and Safety Evaluation of Chinese Medicine, Institution of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China

Zhihua Zhou CAS-Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Tiantian Zuo Tianjin State Key Laboratory of Modern Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin, China

Abbreviations

AACT	Acetyl-CoA C-acetyltransferase
Ach	Acetylcholine
AchE	Acetylcholinesterase
AFLP	Amplified fragment length polymorphism
ANS	Autonomic nervous system
APX	Ascorbate peroxidase
Ara	Arabinose
Aβ	β-amyloid
BUSCO	Benchmarking universal single-copy ortholog
CAS	Cycloartenol synthase
CAT	Catalase
cDNA	Complementary DNA
CEGMA	Core eukaryotic genes mapping approach
COX	Cyclooxygenase
cpDNA	Chloroplast DNA
dCAPS	Derived cleaved amplified polymorphic sequence markers
DD	Dammarenediol-II
DDS	Dammarendiol synthase
DESI-MS	Desorption electrospray ionization mass spectrometry
DMAPP	Dimethylallyl diphosphate
DPPH	2, 2-diphenyl-1-picrylhydrazyl
DPR. Korea	Democratic People's Republic of Korea
EPM	Elevated plus maze
EST	Expression sequences tag
FAD	Fatty acid desaturase
FOS	Fructooligosaccharide
FPP	Farnesyl diphosphate
FPS	Farnesyl diphosphate synthase
Fuc	Fucose
GABA	Gamma-aminobutyrate
Gal	Galactose
GFP	Green fluorescent protein
Glc	Glucose
GluA	Glucuronic acid

GMPGIS	Regionalization Information System of the Ecological Suitability for Medicinal Plants
GOS	Galactooligosaccharide
GP	Genome proportion
GPC	Glycerophosphorylcholine
GPP	Geranyl pyrophosphate
GPX	Glutathione peroxidases
Grx	Glutaredoxin
GT	Glycosyltransferase
HMG-CoA	3-hydroxy-3-methylglutaryl-CoA
HMGR	3-hydroxy-3-methylglutaryl-CoA reductase
HMGS	3-hydroxy-3-methylglutaryl-CoA synthase
HPA	Hypothalamic-pituitary-adrenal
HPGPC	High-performance gel permeation chromatography
HPLC-MS	High-performance liquid chromatography–mass spectrometry
ICTRP	International Clinical Trails Registry Platform
IDI	Isopentenyl-diphosphate delta isomerase
IPP	Isopentenyl diphosphate
IR	Inverted repeat
LAS	Lanosterol synthase
LHCP	Light-harvesting chlorophyll a-b binding proteins
LOX	Lipoxygenase
LSC	Large single copy
LTR	Long tandem repeat
LTR-RT	Long terminal repeat retrotransposon
Man	Mannose
MeJA	Methyl jasmonate
MVA	Mevalonic acid
MVD	Mevalonate diphosphate decarboxylase
MVK	Mevalonate kinase
OA	Oleanolic acid
OAS	Oleanolic acid synthase
OSC	Oxidosqualene cyclase
PD	Parkinson’s disease
PMK	Phosphomevalonate kinase
PPD	Protopanaxadiol
PPT	Protopanaxatriol
PPTS	Protopanaxatriol synthase
PR	Pathogenesis-related protein
Prx	Peroxiredoxin
PSPG	Plant secondary product glycosyltransferase
PVK	Phophomevalonate diphosphate kinase
R.O. Korea	Republic of Korea
RAPD	Random amplified polymorphic DNA
RE	Repetitive element
RFLP	Restriction fragment length polymorphism

Rha	Rhamnose
RNS	Reactive nitrogen species
ROS	Reactive oxygen species
SE	Squalene epoxidase
SNP	Single nucleotide polymorphism
SQE	Squalene epoxidase
SS	Squalene synthase
SSC	Small single copy
TE	Transposable element
TR	Tandem repeat
UDP	Utilize uridine diphosphate
UGT	UDP-glycosyltransferase
UPLC-QTOF-MS	Ultra-performance liquid chromatography quadrupole time-of-flight mass spectrometry
WGD	Whole-genome duplications
Xyl	Xylose