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

Chittaranjan Kole · Hideo Matsumura · Tusar Kanti Behera
Editors

The Bitter Gourd Genome

Springer

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

Tusar Kanti Behera
Division of Vegetable Science
ICAR-Indian Agricultural Research
Institute
New Delhi, India

Hideo Matsumura
Gene Research Center
Shinshu University
Ueda, Japan

ISSN 2199-4781 ISSN 2199-479X (electronic)
Compendium of Plant Genomes
ISBN 978-3-030-15061-7 ISBN 978-3-030-15062-4 (eBook)
https://doi.org/10.1007/978-3-030-15062-4

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 combinations of both technologies, facilitated construction of genetic linkage maps, genetic mapping of genes controlling simply inherited traits, and even gene clusters controlling polygenic traits (QTLs) 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, the 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, the 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 the 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 3 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 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, 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 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 to the Volume

The incidence rate of several deadly diseases, specifically cancer and diabetes, is highly alarming. According to the report of the International Agency for Research on Cancer of World Health Organization, the global cancer burden is estimated to have risen to 18.1 million new cases and 9.6 million deaths in 2018. Worldwide, the total number of people who are alive within 5 years of a cancer diagnosis, called the 5-year prevalence, is estimated to be 43.8 million. According to the projections by the International Diabetes Federation, in 2017, approximately 425 million adults were living with diabetes; by 2045, this will rise to 629 million.

Utilization of medicinal plants and nutraceutical crops is the potential options for alternative and complimentary medicines to mitigate these problems. According to the Zion Market Research, the global herbal supplement market is expected to reach approximately USD 86.74 billion by 2022, growing at a CAGR of around 6.8% between 2017 and 2022. Presently, plant-based drugs contribute 50% to linical drugs. This commercial importance coupled with severe prevalence of the deadly diseases underscores the need for the generation of genetic, genomics, and breeding resources in medicinal plants and functional food crops.

Bitter gourd, also known as African cucumber, ampalaya, balsam pear, balsam apple, bitter apple, and bitter cucumber, is grown traditionally in the tropical and subtropical areas in Asia, South America (Amazon region), East Africa, and the Caribbean as a food vegetable and medicine. This plant contains over 60 phytomedicines potent against 30 diseases. Medicinal properties of bitter melon including antidiabetic, antiviral, antitumor, antileukemic, antibacterial, antihelmintic, antimutagenic, antimycobacterial, antioxidant, antiulcer, anti-inflammatory, hypocholesterolemic, hypotriglyceridemic, hypotensive, immunostimulant, and insecticidal properties have been well documented in research. All parts of this plant, mainly the fruits and the seeds, contain cucurbitacin-B, lycopene, and β-carotene, which are known to have anticancer actions. They also contain charantin and plant insulin that have been clinically demonstrated to have hypoglycemic and anti-hyperglycemic activities and established beneficial effects on diabetes, particularly of type-2. Antioxidant properties of bitter gourd and its traditional use in several countries worldwide including India have been well documented and reviewed in the literature.

Despite immense importance of bitter gourd as a medicinally important vegetable crop and its commercial value, it remained as an "underutilized" or "orphan" crop. Some serious efforts have been started only recently to scientifically demonstrate the medicinal properties of the bioactive compounds in this plant and their mode of actions, and utilization of molecular markers in the evaluation of genetic diversity and elucidation of the genetics of its agro-economic characters. Sequencing of the whole bitter melon genome has been done only in the recent past. Information generated from these molecular genetic and genomic studies will now facilitate breeding of elite varieties with higher fruit yield and content of phytomedicines.

Obviously, there is no book available with compiled information of the botanical descriptions, medicinal properties, genetic and genomic studies, and breeding in this important crop plant. We tried to perform this task by presenting 12 chapters in our book entitled The Bitter Gourd Genome contributed by us and many reputed scientists. Chapter 1 briefs the economic importance of bitter melon as a vegetable crop and also as a medicinal plant and provides glimpses on the works done so far on the areas of genetics, breeding, and genomics. Chapter 2 presents comprehensive information on the botany of the crop under the sections of origin and distribution, taxonomy, morphology, floral biology and mode of reproduction, sex phenology, anatomy, ecology, and economic botany. Chapter 3 deliberates on a variety of bioactive compounds present in bitter gourd including alkaloids, polypeptides, vitamins, and minerals and draws a link of the bioactive compounds to its pharmacological effects like antidiabetic, anticancer, antiviral, anti-inflammatory, analgesic, hypolipidemic, and hypocholesterolemic effects, and provides an insight to understand the mechanism of action. Information on the three gene pools and their potential as a resource of genes useful in breeding has been narrated in Chap. 4. In addition, genetic diversity with regard to morphological characters and content of various phytomedicines has also been delineated. Chapter 5 includes a description of karyotype and chromosomal configurations in this crop and also advanced results from cytomolecular investigations. Cucurbits constitute a unique family including crops with the phenomenon of sex determination. Chapter 6 illuminates the genetic and genomic studies of sex determination in bitter gourd and highlights their potential in elucidating evolution of monoecy and dioecy and more importantly the use of gynoecious lines on crossbreeding. Almost no serious research has been conducted on biotechnology in this crop. However, available information on *in vitro* culture and nanotechnology has been reviewed in Chap. 7. Elaborate discussions on classical genetics and traditional breeding have been covered in Chap. 8 under relevant sections including genetics of agro-economically important qualitative and quantitative traits and details on strategies and tools of conventional breeding. Chapter 9 focuses on various molecular markers and their use in construction of genetic linkage map in bitter gourd. It also deliberates on mapping of many important simply inherited and polygenic traits (QTLs) on these maps. Chapter 10 enumerates on the sequencing of the bitter gourd genome, annotation, and comparison of this genome and its annotation with those in other Cucurbitaceae species that establishes phylogenetical distance of bitter

gourd from other known cucurbit crops and, also the unique properties conferred by the encoding genes, specifically the RIP genes underlying the antitumor or antiviral activities. Since genomics studies have been initiated in bitter gourd much later than many crop plants, only limited works have been done on metabolomics despite of its importance to substantiate the molecular mechanisms of the secondary metabolites such as flavonoids, phenolics, sterols, and terpenoids in this plant in conferring medicinal activities. Chapter 11 briefs the studies on the role of such metabolies, particularly terpenois, on imparting bitterness, medicinal properties, and host responses to pathogens and predators. Finally, a future road has been depicted in Chap. 12 for developing high-density molecular maps, genome sequence with more coverage utilizing advanced sequencing strategies and bioinformatics tools, and ultimately using this information in precise breeding in this crop.

New Delhi, India — Chittaranjan Kole
Ueda, Japan — Hideo Matsumura
New Delhi, India — Tusar Kanti Behera

Contents

Contributors

Vidhu Aeri School of Pharmaceutical Education and Research, New Delhi, India

A. C. Asna Department of Plant Breeding and Genetics, Kerala Agricultural University, Thrissur, Kerala, India

Tusar Kanti Behera Division of Vegetable Sciences, ICAR-Indian Agricultural Research Institute, Pusa, New Delhi, India

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

Shyam Sundar Dey Division of Vegetable Sciences, ICAR-Indian Agricultural Research Institute, Pusa, New Delhi, India

Takeshi Furuhashi Anicom Specialty Medical Institute Inc, Tokyo, Japan

K. K. Gautam ICAR-Indian Institute of Vegetable Research, Varanasi, India

Gograj Singh Jat Division of Vegetable Science, ICAR-Indian Agricultural Research Institute, New Delhi, India

Jiji Joseph Department of Plant Breeding and Genetics, Kerala Agricultural University, Thrissur, Kerala, India

K. Joseph John ICAR-NBPGR Regional Station, Thrissur, Kerala, India

Khalid Mahmood Khawar Department of Crop Science, Faculty of Agriculture, Ankara University, Ankara, Turkey

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

Ricardo A. Lombello Center for Natural and Human Sciences, Federal University of ABC, São Bernardo do Campo, SP, Brazil

Hideo Matsumura Gene Research Center, Shinshu University, Ueda, Nagano, Japan

Sudhakar Pandey ICAR-Indian Institute of Vegetable Research, Varanasi, India

Mamta Pathak Department of Vegetable Science, Punjab Agricultural University, Ludhiana, India

Richa Raj School of Pharmaceutical Education and Research, New Delhi, India

Sevil Saglam Yilmaz Department of Agricultural Biotechnology, Faculty of Agriculture, Kirsehir Ahi Evran University, Kirsehir, Turkey

Naoya Urasaki Okinawa Agricultural Research Center, Itoman, Okinawa, Japan

Abbreviations

½ ×MS medium	Half strength Murashige and Skoog's medium
2,4-D	2, 4 Dichlorophenoxyacetic acid

3 Medicinal

6-BA	6-Benzyladenine
6-BAP	6-Benzylaminopurine
a.a.	Amino acid
ABA	Abscisic acid
ABG-6 medium	½ × MS medium containing 0.5 mg/l BAP
ACC	Aminocyclopropane-1-carboxylic acid
ACS	Aminocyclopropane-1-carboxylic acid synthase
AFLP	Amplified fragment length polymorphism
AKT	Protein kinase
ALT	Alanine aminotransferase
AMPK	Adenosine-5-monophosphate kinase
Apo-A-1	Apolipoprotien A-1
Apo-B	Apolipoprotien B
AP-PCR	Arbitrary primed polymerase chain reaction
AST	Aspartate aminotransferase
ATL	Adult T-cell leukemia
BDMV	Bitter gourd distortion mosaic virus
BLAST	Basic Local Alignment Search Tool
BUSCO	Benchmarking universal single-copy orthologs
Bwa	Burrows–Wheeler Aligner
Cas9	CRISPR-associated protein 9
CCR-B	Cuicurbitacin-B
cDNA	Complementary DNA
CdS	Cadmium sulfide
CDSs	Coding sequences
CHR	Charantin
CMA	Chromomycin A_3
CRISPR	Clustered regularly interspaced short palindromic repeats
CVD	Cardiovascular disease
DAF	DNA amplification fingerprinting
DAMPs	Damage-associated molecular patterns
DAPI	4′,6-Diamidino-2-phenylindole
DaRT	Diversity array technology

DPPH	2,2-Diphenyl-1-picrylhydrazyl
DPPH	1,1-Diphenyl-2-picrylhydrazyl
FIASCO	Functional image analysis software-computational olio
FISH	Fluorescence *in situ* hybridization
FRAP	Ferric reducing ability of plasma
GAE	Gallic acid equivalents
GBS	Genotyping by sequencing
GCA	General combining ability
GC-MS	Gas chromatography–mass spectrometry
GCV	Genotypic coefficient of variation
GD	Genetic distance
GLUT-4	Glucose transport type 4
GP 1	Primary gene pool
GP 2	Secondary gene pool
GP 3	Tertiary gene pool
GSK-3	Glycogen synthase kinase-3
GUS test	β-Glucuronidase test
HDL	High-density lipoprotein
Hep G2	Hepatocellular cancer cell lines
HER 2	Human epidermal growth factor receptor
HIV	Human immunodeficiency virus
HL 60	Human leukemia cells
HR	Hypersensitive response
IBA	Indole 3 butyric acid
IC50	Half maximal inhibitory concentration
IL-1b	Interleukin- 1 beta
IL-6	Interleukin- 6
indel	Insertion and deletion
ISSR	Inter-simple sequence repeat
IU	International unit
JA	Jasmonic acid
LDL	Low-density lipoprotein
LG	Linkage group
LOD	Logarithm of the odds
LPS	Lipopolysaccharide
MAMPs	Microbe-associated molecular patterns
MAP	Mitogen-activated protein
MAP 30	Momordica antiviral protein 30kD
MAS	Marker-assisted selection
matK	Maturase K
MC	*Momordica charantia*
MI	Marker index
MMP-2	Matrix metalloproteinases-2
MMP-9	Matrix metalloproteinases-9
MS medium	Murashige and Skoog medium
NAA	α Naphthalene acetic acid
NCBI	National Center for Biotechnology Information

NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NGS	Next-generation sequencing
NMR	Nuclear magnetic resonance
nptII	Neomycin phosphotransferase II
ORFs	Open reading frames
PAMPs	Pathogen-associated molecular patterns
PAs	Polyamines
PCD	Programmed cell death
PCR	Polymerase chain reaction
PCV	Phenotypic coefficient of variation
PIC	Polymorphic information content
PPR	Pentatricopeptide repeat
PTEN	Phosphatase and tensin homolog
QTL	Quantitative trait locus
QTLs	Quantitative trait loci
QTL-seq	Quantitative trait locus sequencing
RAD	Restriction-associated DNA
RAD-seq	Restriction-associated DNA sequencing
RAD-tags	Restriction-associated DNA tags
RAPD	Random amplified polymorphic DNA
RBG7 medium	½ × MS medium containing 1 mg/l IBA
RFLP	Restriction fragment length polymorphism
RIP	Ribosome inactivating protein
RNA-seq	RNA sequencing
RP	Revolving power
SAR	Systemic acquired resistance
SCA	Specific combining ability
SCAR	Sequence characterized amplified regions
SCT	Seed coat tissue
SNP	Single nucleotide polymorphism
SNVs	Single nucleotide variants
SOD	Superoxide dismutase
Spd	Spermidine
Spm	Spermine
SRAP	Sequence-related amplified polymorphism
SSR	Simple sequence repeat
STS	Sequenced tagged site
t-BHP	Tert-butyl hydroperoxide
TCS	Trichosanthin
TDZ	1-Phenyl-3-(1,2,3-thiadiazol-5-yl)urea or thidiazuron
TLR	Toll-like receptor
TNF-α	Tumor necrosis factor alpha
TPS	Terpene synthase
VLDL	Very low-density lipoprotein
Zn-finger	Zinc finger
α-MMC	α-Momorcharin