
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>

Ilga M. Porth • Amanda R. De la Torre
Editors

The Spruce Genome

 Springer

Editors

Ilga M. Porth
Faculté de foresterie, de géographie
et de géomatique
Université Laval
Quebec, QC, Canada

Amanda R. De la Torre
School of Forestry
Northern Arizona University
Flagstaff, AZ, USA

ISSN 2199-4781 ISSN 2199-479X (electronic)
Compendium of Plant Genomes
ISBN 978-3-030-21000-7 ISBN 978-3-030-21001-4 (eBook)
<https://doi.org/10.1007/978-3-030-21001-4>

© 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

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

The Spruce Genome, an Important Resource to Fundamental Biological Research and Selective Tree Breeding

Main Text

Spruces (*Picea spp.*) are naturally abundant and widely distributed conifer tree species in the Northern hemisphere. Due to their enormous ecological and economic value, management of this important forest genetic resource has focused on conservation and tree improvement. Recently, with the aid of improved sequencing technologies and bioinformatics advances, a draft genome sequence of the 20 Gigabases Norway spruce (*P. abies*) genome was published (Nature 497:581 (2013)). Canadian white spruce hybrid (*P. glauca* × *engelmannii* × *sitchensis*) genome assembly followed in the same year (Bioinformatics 29:1492 (2013)), establishing spruce as a model species in gymnosperm genomics. Continuous efforts to improve the spruce genome assembly are underway, but are challenged by the inherent characteristics of conifer genomes: high amounts of repetitive sequences (introns and transposable elements) and large gene family expansions related to abiotic stress responses, secondary metabolism, and their defense responses against pathogens and herbivory. Because the assembly is still highly fragmented with millions of scaffolds, the generation of ultra-dense genetic maps allows anchoring these scaffolds onto the 12 haploid spruce chromosomes represented by the 12 linkage groups in a spruce genetic map. The generation of RNA-seq data further aids to improve scaffolds. Such data are also particularly valuable in comparative genomics and can highlight the functional divergence between species. Bacterial artificial chromosomes (BACs) sequencing has also served the spruce genomics research community greatly, by (a) unraveling the substantial presence of pseudogenes, (b) supporting the isolation of entire metabolite-biosynthetic genes, (c) facilitating conifer genome comparisons for microsynteny, and last but not least (d) proving indispensable for spruce genome assembly. Some of these BAC sequencing efforts predated the spruce genome sequencing project.

The post-genomic era has seen a surge in genomic applications not only for species amenable to population genomics using whole genome data. In fact, genomics applications using a reduced representation of the massive spruce genome have become very popular (e.g., exome capture sequencing, genotyping-by-sequencing, restriction site-associated DNA sequencing). Throughout this book, we highlight all areas that have been impacted by the acquisition of a high-quality reference genome for spruce. In brief, this volume aims to provide the latest information on (1) status of the genome assembly, (2) detailed insights into whole genome and gene family structure, (3) comprehensive genomic resources available for research, (4) emerging genomics tools for tree improvement programs, (5) genomics related to genetic conservation programs, and (6) functional genomics to improve gene function annotations.

Chapters 1 and 2 focus on the current state of the nuclear and organelle genome assemblies since the first publication of draft genomes and the newest attempts to use whole genome re-sequencing (WGS) data for variant calling. WGS is unprecedented for conifers' complex genomes, where reduced-representation-sequencing-based genotyping has been the state-of-the-art genomic method. By contrast, confident WGS-based variant calling in a population of 1000 individuals for poplar, an angiosperm forest tree species with a 45× smaller genome size, constitutes no major obstacle nowadays. For spruce, however, current challenges regarding such an approach remain. These challenges are highlighted in the respective book chapter on *Picea abies* and potential solutions are extensively discussed. The following two chapters are on repetitive elements, which represent an important 70% fraction of the genome, and retrotransposons are suspected to actually drive spruce genome expansion. The significant differences overall with angiosperm transposable elements dynamics are highlighted; an example also illustrates how BAC sequencing conclusively helps characterize features of a retroelement family important in explaining spruce genome evolutionary dynamics. The epigenomics chapter focuses on the current state of knowledge about epigenetic variation in spruce. One of the chapters devoted to comparative genomics looks at the comparison of nuclear and organelle genomes among spruces, and among spruces and other gymnosperms, focusing on aspects of comparative mapping, and rates of sequence evolution. Another comparative genomics chapter focuses on the sequencing and annotation of a few randomly selected BACs and provides further insights into whole genome evolution comparisons and genome structural features among conifers (spruce versus pine) with regard to genes versus transposons. Spruce genomic resources also have important implications for modern tree selective breeding. A separate chapter is therefore dedicated to genomic selection in white spruce and the increased ability to capture genetic gain by more accurate phenotype prediction models obtained from improved genomic resources. This became possible with the implementation of the genomic pairwise kinship relationship matrix among individuals. This relationship captures the traditional contemporary pedigree (i.e., half- and full-sib family relationships) as well as the historical pedigree through the identification of common DNA variants (SNPs) passed through generations. The following two chapters deal with

local adaptation in spruce, the genetic underpinnings of resistance to drought as well as of cold hardiness. This represents a highlight of the current knowledge in clinal genetic variation in spruces. The last two chapters describe genes and gene families implicated in the formation of terpenes and phenols, the most important secondary compounds in spruce defense. Some of these genes have anti-herbivory and pathogen resistance potential. The book will close with an outlook into emerging fields of research in spruce genomics.

Quebec, Canada
Flagstaff, USA
November 2019

Ilga M. Porth
Amanda R. De la Torre

Contents

1	Sequencing and Assembling the Nuclear and Organelle Genomes of North American Spruces	1
	Inanc Birol and Amanda R. De la Torre	
2	Variant Calling Using Whole Genome Resequencing and Sequence Capture for Population and Evolutionary Genomic Inferences in Norway Spruce (<i>Picea Abies</i>)	9
	Carolina Bernhardsson, Xi Wang, Helena Eklöf, and Pär K. Ingvarsson	
3	Transposable Elements in Spruce	37
	Giovanni Marturano, Camilla Canovi, Federico Rossi, and Andrea Zuccolo	
4	An Intact, But Dormant LTR Retrotransposon Defines a Moderately Sized Family in White Spruce (<i>Picea glauca</i>)	51
	Britta Hamberger, Macaire Man Saint Yuen, Emmanuel Buschiazzo, Claire Cullis, Agnes Yuen, Carol Ritland, Jörg Bohlmann, and Björn Hamberger	
5	The Pliable Genome: Epigenomics of Norway Spruce	65
	Igor Yakovlev, Marcos Viejo, and Carl Gunnar Fossdal	
6	Comparative Genomics of Spruce and Other Gymnosperms	97
	Amanda R. De la Torre	
7	Back to BACs: Conifer Genome Exploration with Bacterial Artificial Chromosomes	107
	Kermit Ritland, Nima Farzaneh, Claire Cullis, Agnes Yuen, Michelle Tang, Joël Fillon, Sarah Chao, Daniel G. Peterson, and Carol Ritland	
8	Genomic Selection in Canadian Spruces	115
	Yousry A. El-Kassaby, Blaise Ratcliffe, Omnia Gamal El-Dien, Shuzhen Sun, Charles Chen, Eduardo P. Cappa, and Ilga M. Porth	

-
- 9 Drought Stress Adaptation in Norway Spruce and Related Genomics Work** 129
Jaroslav Klápště, Jonathan Lecoy,
and María del Rosario García-Gil
- 10 Local Adaptation in the Interior Spruce Hybrid Complex** ... 155
Jonathan Degner
- 11 The Terpene Synthase Gene Family in Norway Spruce** 177
Xue-Mei Yan, Shan-Shan Zhou, Ilga M. Porth,
and Jian-Feng Mao
- 12 Spruce Phenolics: Biosynthesis and Ecological Functions** ... 193
Almuth Hammerbacher, Louwrance P. Wright,
and Jonathan Gershenzon
- 13 Prospects: The Spruce Genome, a Model for Understanding Gymnosperm Evolution and Supporting Tree Improvement Efforts** 215
Ilga M. Porth, Amanda R. De la Torre,
and Yousry A. El-Kassaby