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

Karam B. Singh · Lars G. Kamphuis ·
Matthew N. Nelson
Editors

The Lupin Genome

Springer

Editors
Karam B. Singh
Agriculture and Food
Commonwealth Scientific
and Industrial Research Organisation
Wembley, WA, Australia

Centre for Crop and Disease
Management
Curtin University
Bentley, WA, Australia

The UWA Institute of Agriculture
The University of Western Australia
Perth, WA, Australia

Lars G. Kamphuis
Agriculture and Food
Commonwealth Scientific and Industrial
Research Organisation
Wembley, WA, Australia

Centre for Crop and Disease
Management
Curtin University
Bentley, WA, Australia

The UWA Institute of Agriculture
The University of Western Australia
Perth, WA, Australia

Matthew N. Nelson
Agriculture and Food
Commonwealth Scientific and Industrial
Research Organisation
Wembley, WA, Australia

The UWA Institute of Agriculture
The University of Western Australia
Perth, WA, Australia

Royal Botanic Gardens Kew
Ardingly, West Sussex, UK

ISSN 2199-4781 ISSN 2199-479X (electronic)
Compendium of Plant Genomes
ISBN 978-3-030-21269-8 ISBN 978-3-030-21270-4 (eBook)
https://doi.org/10.1007/978-3-030-21270-4

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 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 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 to the Volume

Introduction to the Lupin Genome

A major development over the last couple of decades has been the development and influence of genomic approaches to help advance our understanding of many crop species; knowledge that in turn can greatly benefit efforts to improve these crops. This book belongs to a book series by Springer called the 'Compendium of Plant Genomes' that describes genomic and related resources in many different crops. It focuses on lupins which are grain legume crops. Legumes, which belong to the Fabaceae (or Leguminosae) family, are widely distributed and form the third-largest land plant family in terms of number of species. From an agricultural point of view, they can occur as both grain crops, also known as pulses, and as fodder crops. Lupins are important ecological 'engineers', able to colonise extremely impoverished soils as well as thrive on low nutrient soils due to their ability to fix atmospheric nitrogen in symbiosis with bacteria and take up phosphorus efficiently from soils.

Lupins belong to the genus *Lupinus* in the Genistoid clade of legumes, which diverged early in Papillionoid legume evolution (Lavin et al. 2005). Lupins are receiving considerable interest recently not only for their value for sustainable farming as a break crop but also as a potential 'superfood' for fighting major health issues around diabetes and obesity. The genus *Lupinus* encompasses around 275 species that are widely distributed geographically, primarily in the Mediterranean region and North and South America, and can be found in a wide range of habitats (Drummond et al. 2012). Only a few lupin species have been domesticated and today the most widely cultivated species are *L. angustifolius* and *L. albus*, while *L. luteus* and *L. mutabilis* are niche crops. Although production has fluctuated over the last 20 years, over a million tonnes are produced every year. In 2017, the largest producers were Australia (1,031,425 t), Poland (168,678 t) and the Russian Federation (161,680 tonnes) (FAO 2017).

This volume on lupin genomics focuses primarily on narrow-leafed lupin (*L. angustifolius* or NLL), which is the main lupin crop primarily grown in Australia. Its genome has been recently sequenced with a focus on the gene-rich space and this has helped lead to the development of new breeding tools for the improvement of this and related lupin crops. This book describes these developments and also has chapters that detail the genomic and related genetic and cytogenetic resources that have been developed for NLL and how they are being used to help advance both fundamental and applied research on NLL in areas ranging from its domestication to syntenic relationships between NLL and other legume crops. Additional chapters report on genomic efforts being undertaken in other lupin crops. A brief outline of the book follows:

Chapter 1 by Dr. Wallace Cowling entitled 'Genetic diversity in narrow-leafed lupin breeding after the domestication bottleneck', and helps set the scene well for the following chapters. Narrow-leafed lupin was not fully domesticated until the 1950s in Australia and Dr. Cowling describes in detail the breeding efforts that led to this achievement and the following efforts to improve the crop. However, the breeding efforts to date have resulted in genome diversity being much lower in domesticated accessions compared with wild relatives, representing a severe domestication bottleneck. Dr. Cowling suggests methods for improving genetic diversity and the potential for long-term genetic gain, including the use of genomic information now available for this crop.

In Chap. 2, entitled 'Ecophysiology and phenology: genetic resources for genetic/genomic improvement of narrow-leafed lupin', Dr. Candy Taylor and colleagues describe the extensive genetic resources available in lupins with a focus on narrow-leafed lupin phenology. They describe how there are around 33,000 accessions of various lupin species that have been accumulated by more than 20 substantially sized and independent gene banks across the globe. They demonstrate how valuable these collections are as resources to breeding programmes to introduce new variation for traits, by focusing on examples related to phenology and in particular flowering time, and how these have benefited crop adaptation in narrow-leafed lupin.

In Chap. 3, entitled 'Overview of genomic resources available for lupins with a focus on narrow-leafed lupin (*Lupinus angustifolius*)', Dr. Karam B. Singh and colleagues provide an overview of the genomic resources available for narrow-leafed lupin with a focus on the current reference genome which underpins many of the other resources. They also describe how the narrow-leafed lupin reference genome has provided valuable insight into narrow-leafed lupin evolution and important information on some of its key plant-microbe interactions. The chapter also touches on some of the genomic resources that are in the pipeline in narrow-leafed and some other lupin species and describes the lupin genome portal, a web-based resource that houses genomic and related information for narrow-leafed lupin.

In Chap. 4, entitled 'Cytomolecular insight into lupin genomes' Dr. Karolina Susek and Dr. Barbara Naganowska summarise a large body of work that has been conducted using cytogenetic approaches in lupins, where again the focus has been on narrow-leafed lupin, which has served as a model for other lupin species. They describe cytogenetic efforts to estimate genome size, identify the number of chromosomes and integrative genetic and cytogenetic mapping in narrow-leafed lupin and discuss how insight into chromosome rearrangements has led to a hypothetical model of lupin karyotype evolution.

Chapter 5, by Dr. Lars G. Kamphuis and colleagues entitled 'Transcriptome resources paving the way for lupin crop improvement' describes the transcriptomic datasets that have been generated for lupin species from expressed sequence tags (EST) libraries through to more recent next generation RNA sequencing libraries. These datasets have been used to generate gene-based molecular markers in lupins, assist with the annotation of the narrow-leafed lupin genome and looked into specific global gene expression studies in different tissue types to address specific research questions around, for example, alkaloid biosynthesis, cluster root formation and organ abscission.

Chapter 6 by Dr. Michał Książkiewicz and Dr. Huaan Yang is entitled 'Molecular marker resources supporting the Australian lupin breeding programme' and provides a detailed overview of the different types of molecular markers that have been used in the Australian and European narrow-leafed lupin breeding programmes. It describes the implementation and accuracy of current molecular markers for domestication traits such as flowering time, seed permeability, pod shattering, alkaloid content, flower colour and disease resistance such as anthracnose caused by *Colletotrichum lupini* and phomopsis stem blight caused by *Diaporthe toxica* and concludes with the opportunities that next generation sequencing has to offer to provide additional molecular markers linked to important traits for lupin crop improvement.

Chapter 7 by Dr. Steven Cannon is entitled 'Chromosomal structure, history, and genomic synteny relationships in *Lupinus*' and capitalises on the genome sequence of narrow-leafed lupin and utilises it to infer the evolutionary history of narrow-leafed lupin relative to other legume species. Using synteny analyses the chapter demonstrates that the ancestor of all lupin species underwent a whole-genome triplication and that chromosome breakages, fusions and independent duplications subsequently led to various chromosome counts in lupin species. It also presents a detailed overview of the online resources generated to view the NLL genome and compare and contrast these to other legumes in various synteny viewers.

The next chapter (Chap. 8) by Dr. Matthew N. Nelson and colleagues entitled 'How have narrow-leafed lupin genomic resources enhanced our understanding of lupin domestication?' focuses on the domestication of lupin species pre- and post- the genomic revolution. It highlights how advances in genetic and genomic technologies have increased our understanding of lupin domestication and how it has led to the identification of key genes that control particular domestication traits such as flowering time and alkaloid content. It also highlights that the domestication process of lupins has led to a significant reduction in genetic diversity in both the Australian and European breeding programs.

Dr. Candy Taylor and colleagues explore the molecular control of time to flowering in narrow-leafed lupin in Chap. 9, which is entitled 'Genomic applications and resources to dissect flowering time control in narrow-leafed lupin'. They describe how modification of phenology was fundamental to the successful adaptation of narrow-leafed lupin to its key growing environments in southern Australia and northern Europe. They go on to recount recent advances in our understanding of the molecular mechanisms underlying these phenology changes, most notably the central role of a *Flowering Locus T* homologue in narrow-leafed lupin.

Chapter 10 by Dr. Paolo Annicchiaricho and colleagues is entitled 'Genetic and genomic resources in white lupin and the application of genomic selection'. Genotyping-by-sequencing technology has enabled cost-effective, accurate and high-density genotyping of white lupin. Two genomic selection approaches were compared and both were able to predict yield, architecture and phenology traits at moderate to high accuracy. The authors then discuss how genomic technology can be applied more broadly to other lupin crops.

In Chap. 11, Dr. Muhammad Munir Iqbal and colleagues review recent advances in 'Genomics of yellow lupin (*Lupinus luteus* L.)'. As a niche crop, yellow lupin has attracted little breeding effort or investment in genomic resources. Recently, the first genetic map for this species was released as well as transcriptomic resources. A genome sequencing project is underway for yellow lupin. The authors discuss how these rapidly developing tools can be used to help plant breeders overcome restraints holding back yellow lupin as a more widely adapted crop.

Finally, in Chap. 12, Dr. Abdelkader Aïnouche and colleagues conducted a detailed analysis of the 'The repetitive content in lupin genomes'. Focusing on four closely related species with striking differences in chromosome number and genome size, they found transposable elements accounted for most of the genome size variation, while many tandem repeats were unique to each species. The authors argue for a centralised resource to house the growing information on the repeat compartment of lupin genomes.

Our hope is that this book will provide a valuable resource to lupin/legume researchers and breeders to understand lupin genomes and a guide on how best to use rapidly developing genomic resources to understand and improve these fascinating legume species.

Wembley, Bentley, Perth, Australia Karam B. Singh
karam.singh@csiro.au
Wembley, Bentley, Perth, Australia Lars G. Kamphuis
lars.kamphuis@csiro.au
Wembley, Perth, Australia Matthew N. Nelson
matthew.nelson@csiro.au

References

Drummond CS, Eastwood RJ, Miotto STS, Hughes CE (2012) Multiple continental radiations and correlates of diversification in *Lupinus* (Leguminosae): testing for key innovation with incomplete taxon sampling. Syst Biol 61:443–460

FAO (2017) FAOSTAT: crop production data. http://www.fao.org/faostat/en/#data/QC: Food and Agriculture Organisation of the United Nations

Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. Syst Biol 54:575–594

Contents

1 **Genetic Diversity in Narrow-Leafed Lupin Breeding After the Domestication Bottleneck** 1
Wallace A. Cowling

2 **Ecophysiology and Phenology: Genetic Resources for Genetic/Genomic Improvement of Narrow-Leafed Lupin** 19
Candy M. Taylor, Lars G. Kamphuis, Wallace A. Cowling, Matthew N. Nelson and Jens D. Berger

3 **Overview of Genomic Resources Available for Lupins with a Focus on Narrow-Leafed Lupin (*Lupinus angustifolius*)** 31
Karam B. Singh, Rhonda C. Foley, Gagan Garg and Lars G. Kamphuis

4 **Cytomolecular Insight Into *Lupinus* Genomes** 45
Karolina Susek and Barbara Naganowska

5 **Transcriptome Resources Paving the Way for Lupin Crop Improvement** 53
Lars G. Kamphuis, Rhonda C. Foley, Karen M. Frick, Gagan Garg and Karam B. Singh

6 **Molecular Marker Resources Supporting the Australian Lupin Breeding Program** 73
Michał Książkiewicz and Hua'an Yang

7 **Chromosomal Structure, History, and Genomic Synteny Relationships in *Lupinus*** 87
Steven B. Cannon

8 **How Have Narrow-Leafed Lupin Genomic Resources Enhanced Our Understanding of Lupin Domestication?** 95
Jemma L. Taylor, Gabriella De Angelis and Matthew N. Nelson

9 **Genomic Applications and Resources to Dissect Flowering Time Control in Narrow-Leafed Lupin** 109
Candy M. Taylor, Lars G. Kamphuis, Wallace A. Cowling, Jens D. Berger and Matthew N. Nelson

10 Genetic and Genomic Resources in White Lupin and the Application of Genomic Selection 139
P. Annicchiarico, N. Nazzicari and B. Ferrari

11 Genomics of Yellow Lupin (*Lupinus luteus* L.) 151
Muhammad Munir Iqbal, William Erskine, Jens D. Berger, Joshua A. Udall and Matthew N. Nelson

12 The Repetitive Content in Lupin Genomes 161
Abdelkader Aïnouche, Aurore Paris, Delphine Giraud, Jean Keller, Pauline Raimondeau, Frédéric Mahé, Pavel Neuman, Petr Novak, Jiri Macas, Malika Aïnouche, Armel Salmon and Guillaume E. Martin

Correction to: Genetic Diversity in Narrow-Leafed Lupin Breeding After the Domestication Bottleneck C1
Wallace A. Cowling