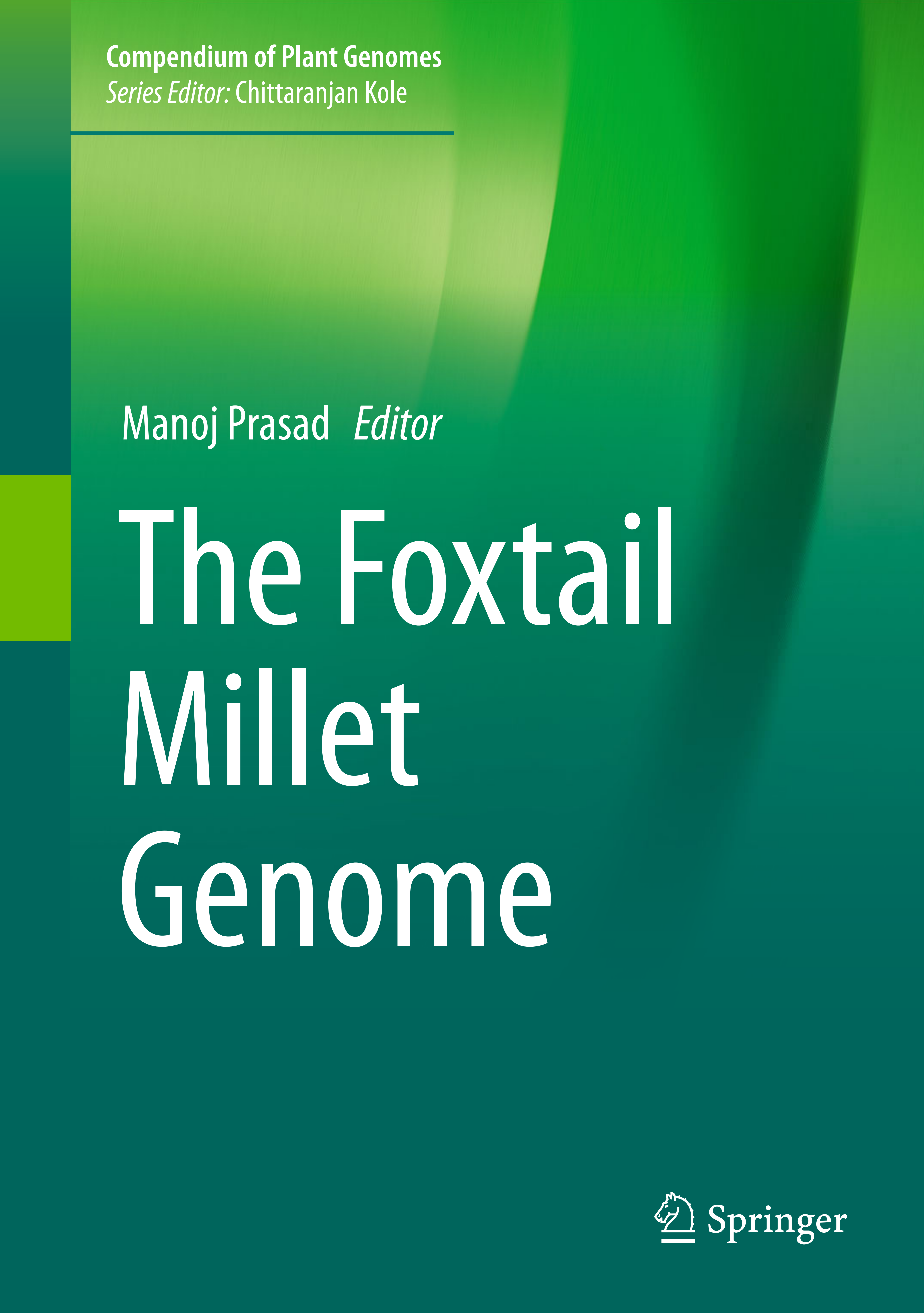

Compendium of Plant Genomes
Series Editor: Chittaranjan Kole

Manoj Prasad *Editor*

The Foxtail Millet Genome

Springer

Compendium of Plant Genomes

Series editor

Prof. Chittaranjan Kole
Raja Ramanna Fellow, Department of Atomic Energy, Government of India,
Kalyani, 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 70 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 Dr. Kole, Series Editor, at ckole2012@gmail.com.

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

Manoj Prasad
Editor

The Foxtail Millet Genome

Springer

Editor
Manoj Prasad
Plant Molecular Biology
National Institute of Plant Genome Research
New Delhi, Delhi
India

ISSN 2199-4781 ISSN 2199-479X (electronic)
Compendium of Plant Genomes
ISBN 978-3-319-65616-8 ISBN 978-3-319-65617-5 (eBook)
DOI 10.1007/978-3-319-65617-5

Library of Congress Control Number: 2017950248

© Springer International Publishing AG 2017
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, express 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.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing 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 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 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 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, 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. 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, 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 are 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 is 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 not only of interest 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, 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 life-time 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, Dr. Christina Eckey and Dr. Jutta Lindenborn in particular, for all their constant and cordial support right from the inception of the idea.

I always had to set aside additional hours to edit books besides 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

Foxtail millet [*Setaria italica* (L.) P. Beauv.] is one of the world's oldest domesticated crops, and it has become a major crop adapted to tropical, sub-tropical, and temperate regions of India, China, and other parts of Asia, North Africa, and the Americas. It has a small diploid genome ($\sim$515 Mb), short growing cycle 0–80 days), self-fertilization, and small morphological stature. Its prolific seed production per plant has made this crop a novel model for grass research. Belonging to the tribe Paniceae in the subfamily Panicoideae, foxtail millet is evolutionarily close to major cereals including sorghum (*Sorghum bicolor*), maize (*Zea mays*), rice (*Oryza sativa*), and candidate biofuel crops such as switchgrass (*Panicum virgatum*), napiergrass (*Pennisetum purpureum*), and pearl millet (*Pennisetum glaucum*). These biofuel crops possess polyploid genomes, long generation times, and large statures, and hence are difficult to study. Therefore, foxtail millet serves as an excellent surrogate genome for studying the genetics and genomics of bioenergy grasses. Efficient biomass productivity and improved water- and nitrogen-use efficiencies are attributable to C_4 photosynthesis, which is assumed to be a prime constituent of exceptional productivity among cereals including maize and sorghum, and bioenergy grasses such as pearl millet and switchgrass. C_4 photosynthesis has evolved as a result of 66 independent plant lineages, and therefore their study is important for understanding the biochemistry, physiology, and molecular biology of C_4 photosynthesis, which could be useful in transferring this system to C_3 species.

Extensive research has been carried out on maize and sugarcane to elucidate the C_4 biology; however, the genome complexity of these crops has resulted in little success. With the availability of reference genomes, haplotype maps, and small genome size, foxtail millet is accentuated as a model system for molecular characterization of C_4 genetics and physiology. In addition, being a staple crop of arid and semi-arid regions of the world, foxtail millet is well known for its tolerance to several abiotic stresses, particularly drought and salinity. This characteristic feature is more important in studying the genetics and genomics of stress tolerance and, therefore, foxtail millet is also considered as a model crop for abiotic stress biology. Recently, the health-benefiting properties of millets have gained importance in nutritional research and, importantly, grains of foxtail millet are reported to be rich in protein, dietary fiber, and energy content, which are three- to fivefold higher than major cereals such as rice, wheat, and maize. Altogether, foxtail millet holds the thrust for establishing climate-resilient agriculture in order to serve nutrient-rich food and fodder to the ever-growing global population.

India tops the list in global millet production; however, pearl millet and finger millet are the top two varieties extensively cultivated in India. Despite the prominent attributes encompassed by foxtail millet, breeding technologies used in this crop are far behind those of pearl millet and finger millet, and other major cereals such as rice and wheat. Foxtail millet has also received very little research attention worldwide, and less effort has been invested towards dissecting the genetic determinants of the prominent traits which are important for improvement of this model species as well as other millets, cereals, and biofuel crops. In this context, the book enumerates the national and international efforts invested in delineating structural, functional, and nutritional genomics of this important crop. The book has 11 chapters describing the general introduction to foxtail millet followed by sequencing efforts, and structural and functional genomics. Chapter 1 provides a general introduction to the crop by outlining its agro-economic importance, origin, distribution, taxonomy, and cytology. Chapter 2 summarizes the sequencing efforts, the outcomes, and the application of sequence information in expediting genomics studies in foxtail millet. Chapter 3 provides interesting information on the impact of transposable elements on distribution pattern and evolution of foxtail millet and its wild progenitor, green foxtail. Chapter 4 summarizes the genetic and genomic resources available in this crop for use in the improvement of foxtail millet per se as well as other related millets, biofuel crops, and cereals. Chapter 5 describes the breeding strategies implemented in foxtail millet and gives an elaborate picture of how the same is being performed in India. Chapters 6 and 7 discuss the association mapping studies and genetic structure analysis of foxtail millet performed using high-throughput molecular markers. Following this, Chap. 8 covers the genetic determinants of abiotic stress tolerance, Chap. 9 the genetic transformation system available in both foxtail millet and green foxtail, Chap. 10 the nutritional potential of this crop and the relevance of nutritional genomics in delineating the health benefiting traits, and Chap. 11 the small RNA biology of foxtail millet. Altogether, the book serves as a primary resource material for researchers, breeders, and students working on millet genomics.

I personally thank Prof. Chittaranjan Kole for giving me the opportunity to edit this interesting book. I also acknowledge the help extended by one of my research scholars, Dr. Muthamilarasan, for his expert assistance in editing and finalizing the chapters.

New Delhi, India Manoj Prasad

Contents

Editor and Contributors

About the Editor

Dr. Manoj Prasad (DOB: 16 December 1970) is working as a Senior Scientist at National Institute of Plant Genome Research (NIPGR), New Delhi, India. He received his Ph.D. degree from Department of Botany, University of Calcutta in 1999. Prior to working at NIPGR, Dr. Prasad served as Alexander von Humboldt Fellow at IPK, Gatersleben, Germany (2000–2003). His research interests include Plant Molecular Genetics and Genomics, and Stress Biology. Dr. Prasad has published more than 120 research papers, reviews, and book chapters in reputed journals. He is serving as Associate Editor in renowned international journals of high repute. He is also a Fellow of three National Academies (INSA, New Delhi, NASI, Allahabad, and NAAS, New Delhi) and has received several awards and honors, including National Bioscience Award of Department of Biotechnology, Government of India, Hiralal Chakraborty Memorial Award, Indian Science Congress Association, India, Alexander von Humboldt Fellow, Germany, and Young Scientist Award and Associateship of NAAS, New Delhi, India.

Contributors

Tirthankar Bandyopadhyay National Institute of Plant Genome Research (NIPGR), New Delhi, India

Charu Lata CSIR-National Botanical Research Institute, Lucknow, India; Academy of Scientific and Innovative Research (AcSIR), New Delhi, India

Kenji Fukunaga Prefectural University of Hiroshima, Shobara, Japan

Gunaseelen Hari-Gowthem National Institute of Plant Genome Research (NIPGR), New Delhi, India

K. Hariprasanna ICAR-Indian Institute of Millets Research, Hyderabad, India

Jinu Jacob ICAR-Indian Institute of Millets Research, Hyderabad, India

Vandana Jaiswal Laboratory of Translational and Evolutionary Genomics, School of Life Science, Jawaharlal Nehru University, New Delhi, India

Mehanathan Muthamilarasan National Institute of Plant Genome Research (NIPGR), New Delhi, India

Parashuram Patroti ICAR-Indian Institute of Millets Research, Hyderabad, India

Manoj Prasad National Institute of Plant Genome Research (NIPGR), New Delhi, India

Radha Shivhare CSIR-National Botanical Research Institute, Rana Pratap Marg, Lucknow 226001, India

Shweta Shweta National Institute of Plant Genome Research (NIPGR), New Delhi, India

Roshan Kumar Singh National Institute of Plant Genome Research (NIPGR), New Delhi, India

Priyanka Sood National Institute of Plant Genome Research (NIPGR), New Delhi, India

K. B. R. S. Visarada ICAR-Indian Institute of Millets Research, Hyderabad, India

Amita Yadav National Institute of Plant Genome Research (NIPGR), New Delhi, India

Chandra Bhan Yadav National Institute of Plant Genome Research (NIPGR), New Delhi, India

Foxtail Millet: An Introduction

1

Roshan Kumar Singh, Mehanathan Muthamilarasan
and Manoj Prasad

Abstract

Foxtail millet (*Setaria italica* L.) is a versatile crop known for being genetically closely related to biofuel grasses, for its C_4 photosynthesis, and for its tolerance to abiotic stresses. These attributes have made this crop a model system and, in view of this, the genome of foxtail millet has been sequenced. Among millets, foxtail millet is the only crop possessing rich genetic and genomic resources, and globally it is the second most cultivated millet next to pearl millet. In the context of its importance in agronomic and research terms, the present chapter summarizes the origin, domestication, phylogeny, and agroeconomic importance of foxtail millet.

1.1 Introduction

The term 'millets' refers to a diverse group of annual cereal crops that characteristically produce small seeds. They include several food, fodder, and biofuel grasses, such as foxtail millet (*Setaria italica*), finger millet (*Elucine coracana*), pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), kodo millet (*Paspalum scorbiculatum*), barnyard millet (*Echinocloa* sp.), etc. (Dwivedi et al. 2012). The major distinctive feature of the millets is their adaptability to cope with adverse agro-ecological conditions such as a semi-dry environment and nutritionally poor soil, the requirement of minimal inputs, and highly nutritious seed content (Lata et al. 2013). Millets require very little water for their cultivation—just around 25–30% of the annual rainfall required by crops such as rice and sugarcane. Thus, millets do not require irrigation and power for their production. In addition, millets also not require any synthetic fertilizers and are completely pest-free crop as none of the millets attracts any pests. Thus, the production of millets is very economical for farmers because of almost nil expenditure on irrigation, fertilizers, and pesticides. Importantly, seeds of most millets can be stored for longer period and are not affected by storage pests. Nutritionally, millets are several times superior to other cereal crops such as rice and wheat. They are rich in minerals

---

R.K. Singh · M. Muthamilarasan · M. Prasad (✉)
National Institute of Plant Genome Research
(NIPGR), Aruna Asaf Ali Marg, New Delhi 110067,
India
e-mail: manoj_prasad@nipgr.ac.in

M. Prasad (ed.), *The Foxtail Millet Genome*, Compendium of Plant Genomes,
DOI 10.1007/978-3-319-65617-5_1