GENOTYPING BY SEQUENCING FOR CROP IMPROVEMENT

EDITED BY
HUMIRA SONAH | VINOD GOYAL | S. M. SHIVARAJ |
RUPESH K. DESHMUKH

WILEY

Table of Contents

List of Tables

List of Illustrations

Genotyping by Sequencing for Crop Improvement

Edited by

Humira Sonah
National Agri-Food Biotechnology Institute
Punjab, India

Vinod Goyal
CCS Haryana Agriculture University
Hisar, India

S.M. Shivaraj
Laval University
Quebec City, QC, Canada

Rupesh K. Deshmukh
National Agri-Food Biotechnology Institute
Punjab, India

WILEY

This edition first published 2022

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Library of Congress Cataloging-in-Publication Data

Names: Sonah, Humira, editor. | Goyal, Vinod, editor. | Shivaraj, S. M., editor. | Deshmukh, Rupesh editor.
Title: Genotyping by sequencing for crop improvement / edited by Humira Sonah, Vinod Goyal, S. M. Shivaraj, Rupesh K. Deshmukh.
Description: First edition. | Hoboken, NJ, USA : John Wiley & Sons, Inc., [2022] | Includes bibliographical references and index.
Identifiers: LCCN 2021046855 (print) | LCCN 2021046856 (ebook) | ISBN 9781119745655 (hardback) | ISBN 9781119745662 (adobe pdf) | ISBN 9781119745679 (epub)
Subjects: LCSH: Genetics–Technique. | Gene mapping. | Genomics. | Plant genomes.
Classification: LCC QK981.45 .G46 2021 (print) | LCC QK981.45 (ebook) | DDC 572.8/62–dc23/eng/20211001
LC record available at https://lccn.loc.gov/2021046855
LC ebook record available at https://lccn.loc.gov/2021046856

Cover Design: Wiley
Cover Image: © Billion Photos/Shutterstock

Dedicated to the two most eminent agricultural scientists of Canada whose work in plant genomics and breeding helped in food security and inspired many young scientists worldwide.

	
Prof. Richard Bélanger Département de phytologie Université Laval, Canada	Prof. François Belzile Département de phytologie Université Laval, Canada

Dr. Humira Sonah
Dr. Vinod Goyal
Dr. S.M. ShivarajDr. Rupesh K. Deshmukh

List of Contributors

Alish Alisha, Department of Gene Expression, Faculty of Biology, Adam Mickiewicz University, Poznań, Poland

Gagandeep Singh Bajwa, Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, Punjab, India

Vitthal T. Barvkar, Department of Botany, Savitribai Phule Pune University, Pune, Maharashtra, India

Shubham Bhardwaj, National Agri-Food Biotechnology Institute (NABI), Mohali, Punjab, India

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

Dharminder Bhatia, Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, Punjab, India

Bharat Char, Mahyco Research Centre, Mahyco Private Limited, Jalna, Maharashtra, India

Viswanathan Chinnusamy, Division of Plant Physiology, ICAR-IARI, New Delhi, India

Shalu Choudhary, Mahyco Research Centre, Mahyco Private Limited, Jalna, Maharashtra, India

Rupesh K. Deshmukh, Agricultural Biotechnology, National Agri-Food Biotechnology Institute (NABI), Mohali, Punjab, India

Vikas Devkar, Department of Plant and Soil Science, Institute of Genomics for Crop Abiotic Stress Tolerance (IGCAST), Texas Tech University, Lubbock, TX, USA

Pallavi Dhiman, Department of Agriculture Biotechnology, National Agri-Food Biotechnology Institute (NABI), Mohali, Punjab, India

Kishor Gaikwad, ICAR – National Institute for Biotechnology, New Delhi, India

Naina Garewal, Department of Biotechnology, Panjab University, Chandigarh, India

Dhananjay Narayanrao Gotarkar, International Rice Research Institute, Los Baños, Philippines

Md Aminul Islam, Department of Botany, Majuli College, Majuli, Assam, India

Priyanka Jain, ICAR – National Institute for Biotechnology, New Delhi, India

Riya Joon, Department of Biotechnology, Panjab University, Chandigarh, India

Swapnil B. Kadam, Department of Botany, Savitribai Phule Pune University, Pune, Maharashtra, India

Ravindra Ramrao Kale, ICAR-Indian Institute of Rice Research, Hyderabad, Telangana, India

Ravneet Kaur, Department of Biotechnology, Panjab University, Chandigarh, India

Suneetha Kota, ICAR-IIRR, Hyderabad, Telangana, India

Amit Kumar, National Agri-Food Biotechnology Institute (NABI), Mohali, Punjab, India

Kuldeep Kumar, ICAR – Indian Institute of Pulses Research, Kanpur, Uttar Pradesh, India

Manish Kumar, Department of Seed Science and Technology, Dr. Yashwant Singh Parmar University of Horticulture and Forestry, Solan, Himachal Pradesh, India

Sandeep Kumar, Xcelris Lab Pvt Ltd., Ahmedabad, Gujarat, India

Virender Kumar, Department of Agriculture Biotechnology, National Agri-Food Biotechnology Institute (NABI), Mohali, Punjab, India

Surbhi Kumawat, Department of Agriculture Biotechnology, National Agri-Food Biotechnology Institute (NABI), Mohali, Punjab, India

Brij Kishore Kushwaha, Department of Molecular Biology and Genetic Engineering, Bihar Agricultural University, Sabour Bhagalpur, Bihar, India

Omkar Maharudra Limbalkar, ICAR-Division of Genetics, Indian Agriculture Research Institute, New Delhi, India

Rushil Mandlik, Department of Agriculture Biotechnology, National Agri-Food Biotechnology Institute (NABI), Mohali, Punjab, India

Venugopal Mikkilineni, Mahyco Research Centre, Mahyco Private Limited, Jalna, Maharashtra, India

Pankaj S. Mundada, Department of Botany, Savitribai Phule Pune University, Pune, Maharashtra, India

Department of Biotechnology, Yashavantrao Chavan Institute of Science, Satara Maharashtra, India

Narender Negi, Department of Fruit Science, ICAR-NBPGR Regional Station, Shimla, Himachal Pradesh, India

Anupama A. Pable, Department of Microbiology, Savitribai Phule Pune University, Maharashtra, India

Gunashri Padalkar, Department of Agriculture Biotechnology, National Agri-Food Biotechnology Institute (NABI), Mohali, Punjab, India

Jayendra Padiya, Mahyco Research Centre, Mahyco Private Limited, Jalna, Maharashtra, India

Arushi Padiyal, Department of Seed Science and Technology, Dr. Yashwant Singh Parmar University of Horticulture and Forestry, Solan, Himachal Pradesh, India

Brajendra Parmar, ICAR-IIRR, Hyderabad, Telangana, India

Gunvant B. Patil, Department of Plant and Soil Science, Institute of Genomics for Crop Abiotic Stress Tolerance (IGCAST), Texas Tech University, Lubbock, TX, USA

Vinaykumar Rachappanavar, Department of Seed Science and Technology, Dr. Yashwant Singh Parmar University of Horticulture and Forestry, Solan, Himachal Pradesh, India

Department of Agriculture, MS Swaminathan School of Agriculture, Shoolini University, Solan, Himachal Pradesh, India

Nitika Rajora, Department of Agriculture Biotechnology, National Agri-Food Biotechnology Institute (NABI), Mohali, Punjab, India

Santosh Rathod, ICAR-IIRR, Hyderabad, Telangana, India

Gaurav Raturi, Department of Agriculture Biotechnology, National Agri-Food Biotechnology Institute (NABI), Mohali, Punjab, India

Rita, ICAR – National Institute for Biotechnology, New Delhi, India

Akshay S. Sakhare, Division of Plant Physiology, ICAR-IARI, New Delhi, India

Swati Saxena, ICAR – National Institute for Biotechnology, New Delhi, India

Senthilkumar Shanmugavel, Crop Improvement Division, ICAR – Sugarcane Breeding Institute, Coimbatore, Tamil Nadu, India

Jitender Kumar Sharma, Department of Agriculture, School of Agriculture, Baddi University of Emerging Sciences & Technology, Baddi, Himachal Pradesh, India

Sandhya Sharma, ICAR – National Institute for Biotechnology, New Delhi, India

Shivani Sharma, National Agri-Food Biotechnology Institute (NABI), Mohali, Punjab, India

Yogesh Sharma, Department of Agriculture Biotechnology, National Agri-Food Biotechnology Institute (NABI), Mohali, Punjab, India

Prashant Raghunath Shingote, Vasantrao Naik College of Agricultural Biotechnology, Dr. Panjabrao Deshmukh Krishi Vidyapeeth, Akola, Maharashtra, India

Harsha Srivastava, ICAR – National Institute for Biotechnology, New Delhi, India

Anuradha Singh, Department of Genomics, ICAR – National Institute on Plant Biotechnology, New Delhi, India

Kashmir Singh, Department of Biotechnology, Panjab University, Chandigarh, India

Manipal Singh, Department of Agriculture Biotechnology, National Agri-Food Biotechnology Institute (NABI), Mohali, Punjab, India

Nisha Singh, Department of Genomics, ICAR – National Institute on Plant Biotechnology, New Delhi, India

Avinash Singode, ICAR – Indian Institute of Millets Research, Hyderabad, Telangana, India

Sweta Sinha, Department of Molecular Biology and Genetic Engineering, Bihar Agricultural University, Sabour Bhagalpur, Bihar, India

Sreeja Sudhakaran, Department of Agriculture Biotechnology, National Agri-Food Biotechnology Institute

(NABI), Mohali, Punjab, India

Lakshmipathy Thalambedu, Crop Improvement Division, ICAR – Sugarcane Breeding Institute, Coimbatore, Tamil Nadu, India

Vandana Thakral, Department of Agriculture Biotechnology, National Agri-Food Biotechnology Institute (NABI), Mohali, Punjab, India

Prathima P. Thirugnanasambandam, Crop Improvement Division, ICAR – Sugarcane Breeding Institute, Coimbatore, Tamil Nadu, India

Anshuman Tiwari, Mahyco Research Centre, Mahyco Private Limited, Jalna, Maharashtra, India

Kishor Tribhuvan, ICAR – Indian Institute of Agricultural Biotechnology, Ranchi, Jharkhand, India

Abhijit Ubale, Mahyco Research Centre, Mahyco Private Limited, Jalna, Maharashtra, India

Sanskriti Vats, Agricultural Biotechnology, National Agri-Food Biotechnology Institute (NABI), Mohali, Punjab, IndiaRegional Centre for Biotechnology, Faridabad, Haryana (NCR Delhi), India

Joshita Vijayan, ICAR – National Institute for Biotechnology, New Delhi, India

Dhiraj Lalji Wasule, Vasantrao Naik College of Agricultural Biotechnology, Dr. Panjabrao Deshmukh Krishi Vidyapeeth, Akola, Maharashtra, India

Himanshu Yadav, Department of Agriculture Biotechnology, National Agri-Food Biotechnology Institute (NABI), Mohali, Punjab, India

Preface

Recent advances in sequencing technology and computational resources have accelerated genomics and translational research in crop science. The technological advances have provided many opportunities in genomics-assisted plant breeding to address issues related to food security. Among the several applications, genotyping-by-sequencing (GBS) technology has evolved as one of the frontier areas facilitating high-throughput plant genotyping. The GBS approaches have proved effective for the utilization in genotyping-based applications like quantitative trait loci (QTL) mapping, genome-wide association study (GWAS), genomic selection (GS), and marker-assisted breeding (MAB). Considering the current affairs in plant breeding, we decided to compile the advances in GBS methods, statistical approaches to analyze the GBS data, and its applications including QTL mapping, GWAS, and GS in crop improvement.

Presently, the food produced around the world is adequate for the existing population. However, the constantly increasing population mounting pressure on a food production system. Hence efficient utilization of technological advances and existing knowledge is essential to enhance food production to match the growing food demand. In this direction, most of the countries around the globe have adopted advanced genomic methodologies to breed superior plant genotypes. Among such technological advances, the high-throughput genotyping using GBS has shown promising results in different crop plants. The GBS has predominantly been used for germplasm evaluation, evolutionary studies, development of dense linkage map, QTL mapping, GWAS, GS, and MAB. The cost-effectiveness

and whole-genome coverage make GBS more reliable than other next-generation sequencing (NGS) techniques.

This book describes advanced molecular markers, high-throughput genotyping platforms, whole-genome resequencing (WGR), QTL mapping using advanced mapping populations, analytical pipelines for the GBS analysis, advances in GWAS, advances in GS, application of GBS, GWAS, and GS in different crop plants. The different marker types including traditional and advanced markers used in plant genotyping have been presented in great detail. DNA extraction directly from seeds without germination can save time and effort. Several modified and crop-specific nondestructive seed DNA extraction protocols have been compiled and presented. Many advanced genotyping platforms are now available which cater to specific research purposes because of the differences in terms of reaction chemistry involved, cost, method of signal detection, and flexibility in the protocols. Such advanced platforms along with their principles have been discussed. The WGR methodology and available resources have been covered in detail. The WGR has emerged as a powerful method to identify genetic variation among individuals. The recent advancement in WGR includes pool-Seq which provides an alternative to individual sequencing and a cost-effective method for GWAS. Compared to biparental populations the multi-parental population provides an opportunity to interrogate multiple alleles and to provide an increased level of recombination and mapping resolution of QTLs. The use of such improved populations in the era of high-throughput genotyping has been presented in one of the chapters. The dedicated section focused on the basic principle of GWAS, the efficiency of different markers, candidate gene identification, meta-GWAS, and statistical methods involved in GWAS analysis has been included. For genetic mapping, and marker-assisted selection, rapid and

quality DNA isolation is mandatory to accelerate the whole process. A focused section about GS has been included which gives an account of the basic concept, advances, applicability, and challenges of GS. Similarly, a separate chapter is included which discusses the analytical pipelines used for GBS data. Application of technologies such as GBS, GWAS, and GS in different crop categories like cereals, pulses, oilseeds, and commercial crops has been discussed in different chapters.

Here, we have tried to compile basic aspects and recent advances in GBS, GWAS, and GS in plant breeding. We believe that the book will be helpful to researchers and scientists to understand and plan future experiments. This book will enable plant scientists to explore GBS application more efficiently for basic research as well as applied aspects in various crops improvement projects.

Editors
Dr. Humira Sonah
Dr. Vinod Goyal
Dr. S. M. ShivarajDr. R. K. Deshmukh

1
Molecular Marker Techniques and Recent Advancements

Dharminder Bhatia and Gagandeep Singh Bajwa
Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, Punjab, India

1.1 Introduction

Plant selection and systematic breeding efforts led to the development of present-day improved cultivars of crop plants. From a historical perspective, increased crop yield is the result of genetic improvement (Fehr 1984). Markers play an important role in the selection of traits of interest. Markers can be morphological, biochemical, or molecular in nature. Morphological markers are visual phenotypic characters such as growth habit of the plant, seed shape, seed color, flower color etc. Biochemical markers are the isozyme-based markers characterized by variation in molecular form of enzyme showing a difference in mobility on an electrophoresis gel. Very few morphological and biochemical markers are available in plants, and they are influenced by developmental stage and environmental factors. Since a large number of economically important traits are quantitative in nature, which are affected by both genetic and environmental factors, the morphological and biochemical markers-based selection of traits may not be much reliable. The subsequent discovery of abundantly available DNA-based markers made possible the selection of almost any trait of interest. DNA-based markers are not affected by the environment. Besides, these markers are highly reproducible across labs and show high polymorphism to distinguish between two genetically different individuals or species.

In the last four decades, DNA-based molecular marker technology has witnessed several advances from low throughput hybridization-based markers to high-throughput sequencing-based markers. These advances have been possible due to critical discoveries such as polymerase chain reaction (PCR) (Mullis et al. 1986), Sanger sequencing method (Sanger et al. 1977), automation of Sanger sequencing (Shendure et al. 2011), next-generation sequencing (NGS) technologies (Mardis 2008), and development of bioinformatics tools. This chapter will briefly discuss different types of molecular markers while particularly focusing on recent developments in molecular marker technologies. These developments have expedited the mapping and cloning of several loci governing important traits, precise trait selection, and transfer into elite germplasm.

1.2 What is a Molecular Marker?

DNA or molecular marker is a fragment of the DNA that is associated with a particular trait in an individual. These molecular markers aid in determining the location of genes that control key traits.

Generally, molecular markers do not represent the gene of interest but act as “flags” or “signs.” Similar to genes, all the molecular markers occupy a specific position within the chromosomes. Molecular markers located close to genes (i.e. tightly linked) are referred to as “gene tags.”

DNA-based molecular markers are the most widely used markers predominantly due to their abundance. They arise from different classes of DNA mutations such as substitution mutations (point mutations), rearrangements (insertions or deletions), or errors in replication of tandemly repeated DNA. These markers are selectively neutral because they are usually located in noncoding regions of DNA. Unlike morphological and