

Huseyin Tombuloglu
Turgay Unver
Guzin Tombuloglu
Khalid Rehman Hakeem *Editors*

Oil Crop Genomics

 Springer

Oil Crop Genomics

Huseyin Tombuloglu • Turgay Unver
Guzin Tombuloglu • Khalid Rehman Hakeem
Editors

Oil Crop Genomics

 Springer

Editors

Huseyin Tombuloglu
Department of Genetics
Institute for Research and Medical
Consultations (IRMC)
Imam Abdulrahman bin Faisal University
(formerly Dammam University)
Dammam, Saudi Arabia

Guzin Tombuloglu
Al Khalidiyah Ash Shamaliyah
Dammam, Saudi Arabia

Turgay Unver
Ficus Biotechnology
Yenimahalle, Ankara, Turkey

Khalid Rehman Hakeem
Department of Biological Sciences
Faculty of Science
King Abdulaziz University
Jeddah, Saudi Arabia

ISBN 978-3-030-70419-3

ISBN 978-3-030-70420-9 (eBook)

<https://doi.org/10.1007/978-3-030-70420-9>

© Springer Nature Switzerland AG 2021

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

***This book is dedicated to our
beloved parents***

Your Lord had decreed, that you worship none save Him, and (that you show) kindness to parents. If one of them or both attain old age with you, say not "Fie" unto them nor repulse them, but speak unto them a gracious word. And lower unto them the wing of submission through mercy and say: My Lord! Have mercy on them both, as they did care for me when I was young.

[Al-Qur'an 17:23-24]

Foreword

Plants are an important source of fats and oils which are essential for the human diet. Genomics of oil biosynthesis in plants has generated great interest in recent years, especially in high oil-bearing plants such as soybean, olive, and oil palm. Considering that, genome-sequencing projects of these plants have been undertaken in recent years with the help of advanced genomics tools like next-generation sequencing. As well, several genome-sequencing projects of oil crops are in progress and many others are on the way. In addition to genome information, advanced genomics approaches such as transcriptomics, genomics-assisted breeding, genome-wide association study (GWAS), genotyping by sequencing (GBS), and CRISPR are improving our understanding of the oil biosynthesis mechanism and breeding strategies for oil extraction and production. However, there is no edited book that covers the genomes and genomics of oil crops, yet.

The book *Oil Crops Genomics* edited by Tombuloglu et al. (Springer) is an interesting and comprehensive scientific work covering recent studies on the genetics and genomics of selected oil crops. The book includes three parts: Part I: Genomes of Oil-Bearing Crops; Part II: Oil Crops Genomics; and Part III: Oil Crop Biotechnology. Part I (Chaps. 1, 2, 3, 4, 5, 6, 7 and 8) covers the genome sequence of oil crops such as soybean, cotton, olive, melon, oil palm, argan tree, and linseed (flax). The second part (Chaps. 9, 10, 11, 12, 13 and 14) concentrates on the genomics of oil crops without genome-sequence information, such as coconut, safflower, mustard, poppy, and *Jatropha*. Part III (Chaps. 15, 16, 17, 18, 19 and 20) represents comprehensive information about the economics of oil crops and the most recent biotechnological methods.

This book combines up-to-date knowledge on genomics and genetics of oil crops. It is well prepared, organized, and represents a valuable source for graduate-level students, instructors, and researchers. Throughout this book, the latest genomics developments and discoveries as well as open problems and future challenges in oil crop genomics are highlighted for future studies. Also, this volume collects the most recent knowledge on oil crop genomics for researchers who study oil crop genome, genomics, biotechnology, pharmacology, medicine nutrition, food industry, or economy.

Ebtesam Abdullah
Al-Suhaimi, PhD
Institute for Research & Medical Consultations
Imam Abdulrahman bin Faisal University
Dammam, Saudi Arabia

Preface

Oil plants are indispensable food sources for human nutrition and health. Especially in recent years, genome analyzes of plants with high oil content such as soybean, olive and palm oil have been carried out intensively. With the determination of genome sequences, it is aimed to better understand the mechanism of oil biosynthesis in plants and to make plants with low oil yield more productive with biotechnological approaches.

The aforementioned developments in oil bearing crops and their novel and advanced studies have not been combined in a single book. While oil production alters depending on the season and years, it causes fluctuations in economical income. Increasing the oil production capacity of wild plants by using genomics-assisted breeding approaches is of great importance to combating this problem. Hence, this volume will be helpful for researchers who study oil crops, their breeding, and biotechnology.

The present book covers genome-sequenced oil crops as well as plants producing important oil metabolites. Throughout this book, the latest genomics developments and discoveries as well as open problems and future challenges in oil crop genomics and their economy are highlighted. We are hopeful that this book shall introduce readers to state-of-the-art developments and trends in oil crop genomic studies.

We are thankful to the contributors for readily accepting our invitation for not only sharing their knowledge and research, but for venerably integrating their expertise in dispersed information from diverse fields in composing the chapters and enduring editorial suggestions to finally produce this venture. We also thank the Springer-International team for their generous cooperation at every stage of the book production.

Dammam, Saudi Arabia
Ankara, Turkey
Dammam, Saudi Arabia
Jeddah, Saudi Arabia

Huseyin Tombuloglu
Turgay Unver
Guzin Tombuloglu
Khalid Rehman Hakeem

Contents

Part I Genomes of Oil-Bearing Crops

1 Soybean Genome	3
Sumayah Alsanie	
2 Overview and Application of Soybean Genomics Study	37
Rong Li, Haifeng Chen, Songli Yuan, and Xinan Zhou	
3 Genetics and Genomics of Cottonseed Oil	53
Jinesh Patel, Edward Lubbers, Neha Kothari, Jenny Koebernick, and Peng Chee	
4 Olive-Tree Genome Sequencing: Towards a Better Understanding of Oil Biosynthesis	75
Mehtap Aydin, Huseyin Tombuloglu, Pilar Hernandez, Gabriel Dorado, and Turgay Unver	
5 Translational Genomics of Cucurbit Oil Seeds	89
Cecilia McGregor and Geoffrey Meru	
6 Genome Sequence of Oil Palm	113
Amal Mahmoud	
7 Argane Genetics and Genomics	123
Hassan Ghazal, Oussama Badad, Houcine Zaid, Tatiana Tatusova, Stacy Pirro, Slimane Khayi, Fatima Gaboun, Kamal Aberkani, Aissam El Finti, Mary Kinsel, Abdelaziz Zahidi, Naima Ait Aabd, Jamila Mouhaddab, Fouad Msanda, Abdellah Idrissi Azami, Rachid Mentag, and Abdelhamid El Mousadik	

8	On “The Most Useful” Oleaginous Seeds: <i>Linum usitatissimum</i> L., A Genomic View with Emphasis on Important Flax Seed Storage Compounds	135
	Lucija Markulin, Yuliia Makhno, Samantha Drouet, Sara Zare, Sumaira Anjum, Duangjai Tungmunnithum, Mohammad R. Sabzalian, Bilal Haider Abbasi, Eric Lainé, Hanna Levchuk, and Christophe Hano	
Part II Oil Crop Genomics		
9	Coconut Genomics	161
	H. D. D. Bandupriya and S. A. C. N. Perera	
10	Complete Chloroplast Genome Sequences of Coconut cv. Kopyor Green Dwarf and Comparative Genome Analysis to Oil Palm, Date Palm, Sago Palm, and Miniature Sugar Palm	189
	Annisa Rahmawati, Hugo Alfried Volkaert, Diny Dinarti, Ismail Maskromo, Andi Nadia Nurul Latifa Hatta, and Sudarsono Sudarsono	
11	Genomics, Phenomics, and Next Breeding Tools for Genetic Improvement of Safflower (<i>Carthamus tinctorius</i> L.)	217
	Abdurrahim Yılmaz, Mehmet Zahit Yeken, Fawad Ali, Muzaffer Barut, Muhammad Azhar Nadeem, Hilal Yılmaz, Muhammad Naeem, Burcu Tarıkahya Hacıoğlu, Yusuf Arslan, Cemal Kurt, Muhammad Aasim, and Faheem Shehzad Baloch	
12	Genomics of Mustard Crops	271
	Umair Riaz, Wajiha Anum, Ghulam Murtaza, Moazzam Jamil, Tayyaba Samreen, Irfan Sohail, Qamar-uz-Zaman, Rashid Iqbal, and Muhammad Ameen	
13	Integrated Omics Analysis of Benzylisoquinoline Alkaloid (BIA) Metabolism in Opium Poppy (<i>Papaver somniferum</i> L.)	291
	Kuaybe Yucebilgili Kurtoglu and Turgay Unver	
14	Transcriptome Analysis in <i>Jatropha</i> During Abiotic Stress Response	317
	Joyce A. Cartagena and Gian Powell B. Marquez	
Part III Oil Crop Biotechnology		
15	Oilseed Crops as the Alternate Source of Omega Fatty Acids: A Paradigm Shift	341
	Sadaf Nazir and Insha Zahoor	
16	Genetic Manipulation for Developing Desired Engineered Oil Crops	353
	Insha Nahvi, Thamer AlShammari, Touseef Amna, and Suriya Rehman	

17 CRISPR Applications in Crops	367
Noha Alqahantani, Bayan Alotaibi, Raghdah Alshumrani, Muruj Bamhrez, Turgay Unver, and Huseyin Tombuloglu	
18 Applications of CRISPR/Cas9 in Oil Crops to Improve Oil Composition	383
Samira Smajlovic, Azra Frkatovic, Hussein Sabit, Huseyin Tombuloglu, and Turgay Unver	
19 Economics of Oil Plants: Demand, Supply, and International Trade	393
Ghulam Mustafa and Asim Iqbal	
20 Production and Trade of Oil Crops, and Their Contribution to the World Economy	415
Dilek Tokel and Bedriye Nazli Erkencioglu	
Index	429

About the Editors



Huseyin Tombuloglu, PhD is an associate professor in the Institute of Research and Medical Consultation (IRMC) at Imam Abdulrahman bin Faisal University, Dammam, Saudi Arabia. He received his BSc degree in 2007 from Istanbul University, Department of Molecular Biology and Genetics, Turkey, and also he studied as an exchange student at the University of Groningen, the Netherlands. Dr. Tombuloglu obtained his MS degree (biology) in 2010 and PhD degree (biotechnology) in 2014. He became an assistant professor in 2014, and associate professor in 2018. He has more

than 15 years of teaching and research experience in genetics, molecular biology, plant genomics, and biotechnology as well as bioinformatics. He is one of the members of the International Olive Genome Sequencing Consortium. Currently, he is in the research team member of the International Garlic Genome Sequencing Consortium (IGGS) (<http://garlicgenome.org/>). His current research is focused on genome sequencing of plants, data analysis, proteomics, and nanoparticle-plant interaction.



Turgay Unver, PhD received his PhD degree from Middle East Technical University (METU; Turkey) in 2008. He became an assistant professor at Cankiri Karatekin University in 2009, and obtained his associated professorship in 2011. His main research areas involve genome and transcriptome analyses, including microRNA. He has published >50 indexed articles with more than 4000 citations. In 2012, he was awarded the TWAS prize. He was selected as outstanding young scientist by the Turkish Academy of Science in 2013. Prof. Unver is currently running two biotechnology companies and bearing editorial duties for *Genomics* (Elsevier), *BMC Genomics*, and *Plos One*.



Guzlin Tombuloglu, PhD received her MS degree (biology) in 2008 and PhD degree (biotechnology) in 2014. She has experience in transcriptome sequencing, plant abiotic stress tolerance, and molecular biology of plants. During her PhD, she studied transcriptomics identification of boron tolerance mechanism in barley. She has experienced several projects on abiotic stress, plant stress responses, boron toxicity, and transcriptomics. She has given several courses on teaching genetics, molecular biology, and biotechnology education for more than 15 years. She also worked as chairman of the Pathology Laboratory Techniques Programme and assistant manager at the Vocational School of Medical Sciences at university level.



Khalid Rehman Hakeem, PhD is a professor at King Abdulaziz University, Jeddah, Saudi Arabia. After completing his doctorate (botany; specialization in plant eco-physiology and molecular biology) from Jamia Hamdard, New Delhi, India, in 2011, he worked as assistant professor at the University of Kashmir, Srinagar, for a short period. Later, he joined Universiti Putra Malaysia, Selangor, Malaysia, and worked there as postdoctorate fellow in 2012 and fellow researcher (associate professor) from 2013 to 2016. Dr. Hakeem has more than 10 years of teaching and research experience in plant eco-physiology, biotechnology and molecular biology, medicinal plant research, plant-microbe-soil interactions, as well as in environmental studies. He is the recipient of several fellowships at both national and international levels. He has also served as a visiting scientist at Jinan University, Guangzhou, China. Currently, he is involved with a number of international research projects with different government organizations. So far, Dr. Hakeem has authored and edited more

experience in plant eco-physiology, biotechnology and molecular biology, medicinal plant research, plant-microbe-soil interactions, as well as in environmental studies. He is the recipient of several fellowships at both national and international levels. He has also served as a visiting scientist at Jinan University, Guangzhou, China. Currently, he is involved with a number of international research projects with different government organizations. So far, Dr. Hakeem has authored and edited more

than 70 books with international publishers, including Springer Nature, Academic Press (Elsevier), and CRC Press. He also has to his credit more than 140 research publications in peer-reviewed international journals and 60 book chapters in edited volumes with international publishers. At present, Dr. Hakeem serves as an editorial board member and reviewer of several high-impact international scientific journals from Elsevier, Springer Nature, Taylor and Francis, Cambridge, and John Wiley Publishers. He is included in the advisory board of Cambridge Scholars Publishing, UK. Dr. Hakeem is also a fellow of Plantae group of the American Society of Plant Biologists; member of the World Academy of Sciences; member of the International Society for Development and Sustainability, Japan; and member of Asian Federation of Biotechnology, Korea. Dr. Hakeem has been listed in Marquis Who's Who in the World, between 2014 and 2020. Currently, Dr. Hakeem is engaged in studying the plant processes at eco-physiological as well as molecular levels.

Contributors

Naima Ait Aabd CRRRA-Agadir, National Institute for Agricultural Research (INRA), Rabat, Morocco

Muhammad Aasim Faculty of Agricultural Sciences and Technologies, Sivas University of Science and Technology, Sivas, Turkey

Bilal Haider Abbasi Department of Biotechnology, Quaid-i-Azam University, Islamabad, Pakistan

Kamal Aberkani Polydisciplinary Faculty of Nador, University Mohammed Premier, Oujda, Morocco

Fawad Ali Department of Plant Sciences, Quaid-I-Azam University, Islamabad, Pakistan

Bayan Alotaibi Department of Genetics Research, Institute for Research and Medical Consultations (IRMC), Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

Noha Alqahantani Department of Genetics Research, Institute for Research and Medical Consultations (IRMC), Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

Sumayah Alsanie Biology Department, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

Thamer AlShammari Department of Genetic Research, Institute for Research & Medical Consultations (IRMC), Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

Raghdah Alshumrani Department of Genetics Research, Institute for Research and Medical Consultations (IRMC), Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

Muhammad Ameen Institute of Soil and Environmental Sciences, University of Agriculture, Faisalabad, Pakistan

Touseef Amna Department of Biology & Biotechnology, Faculty of Science, Albaha University, Albaha, Saudi Arabia

Sumaira Anjum Department of Biotechnology, Kinnaird College for Women, Lahore, Pakistan

Wajiha Anum Department of Agronomy, Regional Agricultural Research Institute, Bahawalpur, Pakistan

Yusuf Arslan Department of Field Crops, Faculty of Agriculture and Natural Sciences, Bolu Abant Izzet Baysal University, Bolu, Turkey

Mehtap Aydin Genetics and Bioengineering Department, Yeditepe University, Istanbul, Turkey

Abdellah Idrissi Azami School of Medicine and Pharmacy, University Mohammed V, Rabat, Morocco

Oussama Badad Faculty of Sciences, University Mohammed V, Rabat, Morocco

Faheem Shehzad Baloch Faculty of Agricultural Sciences and Technologies, Sivas University of Science and Technology, Sivas, Turkey

Muruj Bamhrez Department of Genetics Research, Institute for Research and Medical Consultations (IRMC), Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

H. D. D. Bandupriya Department of Plant Sciences, Faculty of Science, University of Colombo, Colombo, Sri Lanka

Muzaffer Barut Department of Field Crops, Faculty of Agriculture, Çukurova University, Adana, Turkey

Joyce A. Cartagena Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Japan

Peng Chee Cotton Molecular Breeding Laboratory, University of Georgia – Tifton Campus, Tifton, GA, USA

Haifeng Chen Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs of PRC, Oil Crops Research Institute of Chinese Academy of Agriculture Sciences, Wuhan, China

Diny Dinarti Plant Molecular Biology Lab., Department of Agronomy and Horticulture, Faculty of Agriculture, Bogor Agricultural University, Bogor, West Java, Indonesia

Gabriel Dorado Dep. Bioquímica y Biología Molecular, Campus Rabanales C6-1-E17, Campus de Excelencia Internacional Agroalimentario (ceiA3), Universidad de Córdoba, Córdoba, Spain

Samantha Drouet Laboratoire de Biologie des Ligneux et des Grandes Cultures, INRAE USC1328, Orleans University, Orléans Cedex 2, France

Aissam El Finti Laboratory of Biotechnology and Valorization of Natural Resources (LBVRN), Faculty of Sciences, University Ibn Zohr, Agadir, Morocco

Azra Frkatovic Genos Glycoscience Research Laboratory, Zagreb, Croatia

Fatima Gaboun CRRA-Rabat, National Institute for Agricultural Research (INRA), Rabat, Morocco

Bedriye Nazli Genc Department of Medical Pathology, Faculty of Medicine, Istinie University, Istanbul, Turkey

Hassan Ghazal National Center for Scientific and Technological Research (CNRST), Rabat, Morocco

Christophe Hano Laboratoire de Biologie des Ligneux et des Grandes Cultures, INRAE USC1328, Orleans University, Orléans Cedex 2, France

Andi Nadia Nurul Latifa Hatta Plant Molecular Biology Lab., Department of Agronomy and Horticulture, Faculty of Agriculture, Bogor Agricultural University, Bogor, West Java, Indonesia

Rashid Iqbal Department of Agronomy, Islamia University of Bahawalpur, Bahawalpur, Pakistan

Asim Iqbal Department of Economics and Business Administration, Division of Arts and Social Sciences, University of Education, Lahore, Pakistan

Moazzam Jamil Department of Soil Science, Islamia University of Bahawalpur, Bahawalpur, Pakistan

Slimane Khayi CRRA-Rabat, National Institute for Agricultural Research (INRA), Rabat, Morocco

Mary Kinsel Department of Chemistry & Biochemistry, Southern Illinois University in Carbondale, Carbondale, IL, USA

Jenny Koebernick Department of Crop, Soil and Environmental Sciences, Auburn University, Auburn, AL, USA

Neha Kothari Fiber Quality Research, Cotton Incorporated, Cary, NC, USA

Cemal Kurt Department of Field Crops, Faculty of Agriculture, Çukurova University, Adana, Turkey

Kuaybe Yucebilgili Kurtoglu Faculty of Science, Department of Molecular Biology and Genetics, Istanbul Medeniyet University, Istanbul, Turkey

Eric Lainé Laboratoire de Biologie des Ligneux et des Grandes Cultures, INRAE USC1328, Orleans University, Orléans Cedex 2, France

Hanna Levchuk Flax Breeding Lab, Institute of Oilseed Crops of the National Academy of Agricultural Sciences of Ukraine, Ukraine

Rong Li Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs of PRC, Oil Crops Research Institute of Chinese Academy of Agriculture Sciences, Wuhan, China

Edward Lubbers Cotton Molecular Breeding Laboratory, University of Georgia – Tifton Campus, Tifton, GA, USA

Amal Mahmoud Department of Biology, College of Science, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

Yuliia Makhno Flax Breeding Lab, Institute of Oilseed Crops of the National Academy of Agricultural Sciences of Ukraine, Ukraine

Lucija Markulin Laboratoire de Biologie des Ligneux et des Grandes Cultures, INRAE USC1328, Orleans University, Orléans Cedex 2, France

Gian Powell B. Marquez College of Global Liberal Arts, Ritsumeikan University, Osaka, Japan

Ismail Maskromo Indonesian Palms Research Institute, Agency for Agricultural Research and Development, Manado, Indonesia

Cecilia McGregor Institute of Plant Breeding, Genetics & Genomics and Department of Horticulture, University of Georgia, Athens, GA, USA

Rachid Mentag CRRA-Rabat, National Institute for Agricultural Research (INRA), Rabat, Morocco

Geoffrey Meru Department of Horticulture, Tropical Research & Education Center, University of Florida, IFAS, Homestead, FL, USA

Jamila Mouhaddab Laboratory of Biotechnology and Valorization of Natural Resources (LBVRN), Faculty of Sciences, University Ibn Zohr, Agadir, Morocco

Abdelhamid El Mousadik Laboratory of Biotechnology and Valorization of Natural Resources (LBVRN), Faculty of Sciences, University Ibn Zohr, Agadir, Morocco

Fouad Msanda Laboratory of Biotechnology and Valorization of Natural Resources (LBVRN), Faculty of Sciences, University Ibn Zohr, Agadir, Morocco

Ghulam Murtaza Institute of Soil and Environmental Sciences, University of Agriculture, Faisalabad, Pakistan

Ghulam Mustafa Department of Economics and Business Administration, Division of Arts and Social Sciences, University of Education, Lahore, Pakistan

Muhammad Azhar Nadeem Faculty of Agricultural Sciences and Technologies, Sivas University of Science and Technology, Sivas, Turkey

Muhammad Naeem Department of Plant Breeding and Genetics, Faculty of Agriculture and Environmental Sciences, The Islamia University of Bahawalpur, Bahawalpur, Punjab, Pakistan

Insha Nahvi Department of Basic Sciences, Preparatory Year Deanship, King Faisal University, Hofuf, Saudi Arabia

Sadaf Nazir Department of Food Technology, Institute of Engineering & Technology, Bundelkhand University, Jhansi, Uttar Pradesh, India

Jinesh Patel Department of Crop, Soil and Environmental Sciences, Auburn University, Auburn, AL, USA

S. A. C. N. Perera Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka

Stacy Pirro Iridian Genome, Bethesda, MD, USA

Qamar-uz-Zaman Department of Environmental Sciences, University of Lahore, Lahore, Pakistan

Annisaa Rahmawati Plant Molecular Biology Lab., Department of Agronomy and Horticulture, Faculty of Agriculture, Bogor Agricultural University, Bogor, West Java, Indonesia

Suriya Rehman Department of Epidemic Diseases Research, Institute for Research & Medical Consultations (IRMC), Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

Umair Riaz Soil and Water Testing Laboratory for Research, Agriculture Department, Government of Punjab, Bahawalpur, Pakistan

Hussein Sabit Department of Genetics Research, Institute for Research and Medical Consultations (IRMC), Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

Mohammad R. Sabzalian Department of Agronomy and Plant Breeding, College of Agriculture, Isfahan University of Technology (IUT), Isfahan, Iran

Tayyaba Samreen Institute of Soil and Environmental Sciences, University of Agriculture, Faisalabad, Pakistan

Samira Smajlovic Faculty of Science, Department of Biology, University of Zagreb, Zagreb, Croatia

Irfan Sohail Institute of Soil and Environmental Sciences, University of Agriculture, Faisalabad, Pakistan

Sudarsono Sudarsono Plant Molecular Biology Lab., Department of Agronomy and Horticulture, Faculty of Agriculture, Bogor Agricultural University, Bogor, West Java, Indonesia

Burcu Tarıkahya Hacıoğlu Department of Biology, Faculty of Science, Hacettepe University, Ankara, Turkey

Tatiana Tatusova National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD, USA

Dilek Tokel Department of Economics, Faculty of Economics, Marmara University, Istanbul, Turkey

Huseyin Tombuloglu Department of Genetics Research, Institute for Research and Medical Consultations (IRMC), Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

Duangjai Tungmunnithum Department of Pharmaceutical Botany, Faculty of Pharmacy, Mahidol University, Bangkok, Thailand

Turgay Unver Ficus Biotechnology, Ankara, Turkey

Hugo Alfried Volkaert Center for Agricultural Biotechnology, Kasetsart University, Nakhon Pathom, Thailand

Abdurrahim Yılmaz Department of Field Crops, Faculty of Agriculture and Natural Sciences, Bolu Abant Izzet Baysal University, Bolu, Turkey

Hilal Yılmaz Department of Plant and Animal Production, Izmit Vocational School, Kocaeli University, Kocaeli, Turkey

Mehmet Zahit Yeken Department of Field Crops, Faculty of Agriculture and Natural Sciences, Bolu Abant Izzet Baysal University, Bolu, Turkey

Songli Yuan Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs of PRC, Oil Crops Research Institute of Chinese Academy of Agriculture Sciences, Wuhan, China

Abdelaziz Zahidi Laboratory of Biotechnology and Valorization of Natural Resources (LBVRN), Faculty of Sciences, University Ibn Zohr, Agadir, Morocco

Insha Zahoor Department of Neurology, Henry Ford Hospital, Detroit, MI, USA

Houcine Zaid Faculty of Sciences, University Mohammed V, Rabat, Morocco

Sara Zare Department of Agronomy and Plant Breeding, College of Agriculture, Isfahan University of Technology (IUT), Isfahan, Iran

Xinan Zhou Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs of PRC, Oil Crops Research Institute of Chinese Academy of Agriculture Sciences, Wuhan, China

Part I
Genomes of Oil-Bearing Crops

Chapter 1

Soybean Genome



Sumayah Alsanie

Contents

1.1	Introduction.....	3
1.2	Soybean Taxonomy and Morphology.....	6
1.2.1	Soybean Morphology.....	8
1.3	Soybean Cytogenetics.....	11
1.4	Soybean Genome.....	13
1.4.1	Soybean Genome Assembly.....	13
1.4.2	The Sequence of Soybean's Chloroplast and Mitochondria Genome.....	16
1.4.3	Website Links.....	17
1.5	Soybean Genetic Resources (Germplasm).....	18
1.5.1	A List of Some Germplasm Resources for Soybeans.....	24
1.6	Economic Importance of Soybean.....	24
	References.....	28

1.1 Introduction

Glycine max (L.) Merr. (soybean) is a subtropical plant native to southeastern Asia. It has been a part of the human diet in Asian countries for a very long period. European countries and the USA were introduced to this dietary staple comparatively very recently, the 1700s and 1800s, respectively (Joy et al. 1998). It is the widely grown crop in the world and the fourth most important harvested and produced crop (Schmutz et al. 2010).

According to the recent statistics, soybean constitutes 59% of oilseed production globally in 2019 (Fig. 1.1), and the USA and Brazil are the top producers of the soybean (Soystats 2020).

The soybean is native to East Asia. The widely cultivated soya bean *Glycine max* belongs to subgenus *Soja* of the genus *Glycine* (which has at least 25 perennial species). The other wild soybean, *Glycine soja* Sieb and Zucc, also belongs to the subgenus *Soja* of the genus *Glycine*. *Glycine soja* and *Glycine max* both are annual plants. The widely cultivated, modern soybean can no longer be traced to the

S. Alsanie (✉)

Biology Department, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia
e-mail: sualsanea@iau.edu.sa

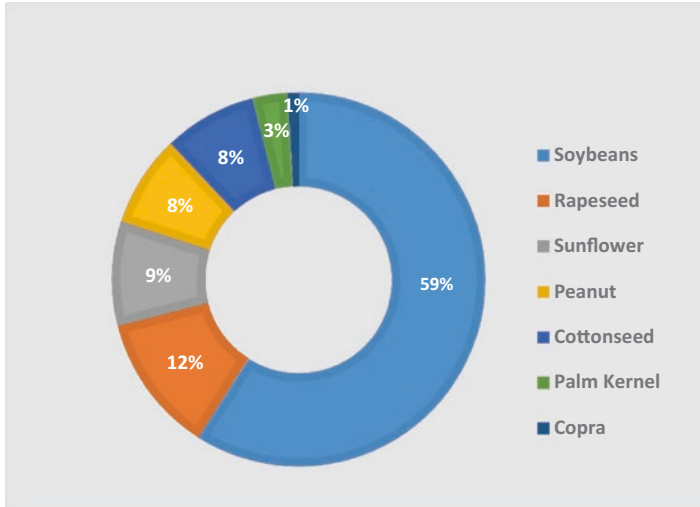


Fig. 1.1 Percentage of world oilseed production (2019). (Data from soystats2020.com, United States Department of Agriculture (USDA), foreign agricultural service (FAS))

wild-growing species (Shekhar et al. 2016). When those two species are crossed together, they produce fertile hybrids, which concludes that they have similar genomes (Singh and Hymowitz 1988). While the 26 wild perennial species originate in Australia, they differ morphologically, cytologically, and genomically from species that belong to subgenus *Soja* (Chung and Singh 2008).

Soybean is an economically important crop as one of the primary sources of oil and protein. Soybean is solvent-extracted for oil, and the remaining part is “toasted.” The soy meal thus produced contains 50% protein, which is then used as animal feed (e.g., chicken, hog, turkey). Soybean is also an important part of the several processed foods. It gained importance during World War II as an equivalent alternative for protein foods and as a source of edible oil (Shurtleff and Aoyagi 2009).

The soybean crop requires a climate with hot summers (temperatures range 20–30 °C). The temperature outside of this range results in stunted growth. Soybean growth is optimum in moist soils with rich organic content but also grows in other types of soil. Soybean roots have nodules for nitrogen-fixing bacterium like other legumes that perform nitrogen fixation. The crop is favored for rotational cropping for its nitrogen fixation.

The soybean crop is affected by the day length – longer day results in taller plants, more nodes, and delayed flowering. Conversely, shorter days result in early flowering. It requires 100–130 days to grow, and there is no vegetative growth during the yield formation.

The soybean yield is highly dependent on water availability, fertilization, seed variety, and row spacing. Improved varieties can yield as high as 2.5 and 3.5 ton/ha, while under rain-fed conditions a soybean yield is around 1.5–2.5 ton/ha. Although irrigation doesn’t affect the oil and protein content of the seed significantly, a slight

increase in protein and a small decrease in oil content may occur (Food and Agriculture Organization of the United Nations (FAO) 2020).

Soybean has high (38–45%) protein and high (approximately 20%) oil content. After the oil extraction, the material remaining is called soybean meal, or soy meal, which contains 50% soy protein. Soybean constitutes major part of the livestock feed worldwide. About 97% of livestock feed is produced from soybean, which serves as protein source (Kansas Soybean Commission, <https://kansassoybeans.org/about-the-checkoff/animal-ag/>). Some of it is also used in some dog foods (Menniti et al. 2014). In the European Union, 60% of the protein fed to livestock comes from the soy meal (Heuzé et al. 2020).

Protein content in the soy flour is 49% and the fiber content is 18 gm. Soy flour is gluten-free. Also, protein and other nutrient content in the soy flour are higher compared to wheat flour (Nutrition data, <https://nutritiondata.self.com/facts>).

Apart from the livestock feed, soybean products have found a way in the human diet and have gained an important part. Commonly used products include soybean oil, soy sauce, soy milk, tofu, soy flour, tempeh, textured vegetable protein, and soy lecithin. In Japan, edamame is a minimally processed dish of soybean. Soy flour is used in sauces, baked food, and frying (to reduce oil absorption). The soy flour imparts tenderness, moistness, and texture to baked food; hence it is used in baked foods. Soybean has been part of food traditionally in the form of soy milk and tofu.

Soybeans can be processed into various forms to substitute for a variety of food items having a comparable texture and appearance. Some of the examples of dairy products that similar soybean products can substitute are milk, margarine, ice cream, yogurt, and cheese. It can also be an alternative to meat as in veggie burgers. It can also be a cost-effective replacement for meat and poultry products.

However, the digestible calcium content in soy milk is not significant. Hence, many calcium-enriched products of soy milk are also manufactured. Though the protein quality in the soy products is roughly equivalent to the natural products, the products can be fortified with vitamin and mineral to enrich nutrient content to the equivalent level. The soy-based meat substitute has been in use to replace the ground beef cost-effectively and without compromising on the nutrition (Van Wyk 2005; Rizzo and Baroni 2018; North Carolina soybean production association (ncsoy.org)). Soy nut butter is another product that can substitute for peanut butter (Shurtleff and Aoyagi 2012).

Soy-based infant formula is another usage of soybean. It is advantageous for infants who need breastfeeding or wanted a vegan diet. It is a good substitute for infants allergic to pasteurized cow milk proteins. Different powdered, ready-to-feed forms are available in the market (American Academy of Pediatrics 1998; Rizzo and Baroni 2018).

Considering the various uses of the soybean as food, demand for soybean is increasingly high (Soystats 2020). However, viral and pests infections can reduce production substantially.

On a global scale, approximately 191 million hectares of genetically engineered crops were planted in 26 countries in 2017 (International Service for the Acquisition of Agri-biotech Applications (ISAAA) 2018).

Soybeans have been genetically modified in recent few years for better yield and other advantages. Several products are made from the genetically modified soybeans. The genetic modification started in 1996 when Monsanto Company introduced a genetically modified variety by substituting *Agrobacterium* sp. (strain CP4) gene EPSP (5-enolpyruvyl-shikimic acid-3-phosphate) synthase (Padgett et al. 1995; Duke and Cerdeira 2010). It was a herbicide-tolerant variety. Also, studies were conducted to develop *CpMMV* (Cowpea Mild Mottle Virus)-resistant variety.

In Brazil, the adaption of genetically modified soya bean increased from about 34% of soya bean acres in 2004 to 96% in 2019 (Soystats 2020). Though the cultivation of genetically modified soybean has increased exponentially, it has affected the export of soya bean to the European Union as a considerable number of suppliers and consumers are reluctant to use genetically modified products like food or animal feed. Hence they require extensive certification before export.

An extensive genetic diversity is observed in the soybean due to its predisposition to adapt to a variety of environmental conditions (Li et al. 2010; Zhou et al. 2015). Breeders choose the best varieties to improve soybean cultivation and ultimately the soybean yield. Therefore, in order to advance the marker-assisted breeding programs of soybean, knowledge of fundamental genetics of agronomical traits is essential (Wang et al. 2009). In the future, plant breeding in combination with recent advances in genomics will drive the expectations for soybean improvement (Palmer and Hymowitz 2016).

1.2 Soybean Taxonomy and Morphology

The taxonomy of cultivated soybean is presented below:

Order	Fabales
Family	Fabaceae (Leguminosae)
Subfamily	Papilionoideae (Faboideae)
Tribe	Phaseoleae
Subtribe	Glycininae
Genus	<i>Glycine</i> Willd
Subgenus	<i>Soja</i> (Moench) F. J. Herm.
Species	<i>G. max</i>
Botanical name	<i>Glycine max</i> (L.) Merr
Synonyms	<i>G. gracilis</i> , <i>G. Soja</i>
Common name	Soybean, soya bean

(Joy et al. 1998; Ratnaparkhe et al. 2011)

The Fabaceae or Leguminosae family is a large family of flowering plants which includes around 20,000 species (Lewis et al. 2005). The three subfamilies of Fabaceae are Caesalpinieae, Mimosoideae, and Papilionoideae. The subfamily

Papilionoideae includes the species grown as crops such as soybean, pea, common bean, mung bean, cowpea, etc. (Lewis et al. 2005; Bruneau et al. 2008).

Genus *Glycine* Willd has approximately 25 perennial species (Greek word *glykys* = sweet). The term *Glycine* was coined by Linnaeus (Hymowitz and Newell 1981). The taxonomy tree was first arranged by Bentham (1864, 1865) (Table 1.1) who divided the genus *Glycine* into three sections: *Johnia*, *Soja*, and *Leptocytamus*. The cultivated soybean species was included in section *Soja*.

Table 1.1 Development of *Glycine* taxonomy according to different researchers

Bentham (1864, 1865)	Hermann (1962)	Verdcourt (1966, 1970)	Newell And Hymowitz (1980)	Current
Section <i>Johnia</i> <i>G. javanica</i> –	Subgenus <i>Glycine</i> <i>G. javanica</i> <i>G. petitiانا</i>	Subgenus <i>Bracteata</i> <i>G. wightii</i> = <i>Neonotonia</i>	–	–
Section <i>Soja</i>	Subgenus <i>Soja</i>	Subgenus <i>Soja</i>	Subgenus <i>Soja</i> (Moench) F. J. Herm.	Subgenus <i>Soja</i> (Moench) F. J. Herm.
<i>G. soja</i> (cultivated)	<i>G. ussuriensis</i>	<i>G. soja</i> (previously <i>G. ussuriensis</i>)	<i>G. soja</i> Sieb. and Zucc.	<i>G. soja</i> Sieb. and Zucc.
<i>G. hedyaroides</i> = <i>Ophrestia</i>	–	–	–	–
<i>G. pentaphylla</i> = <i>Ophrestia</i>	–	–	–	–
<i>G. lyalli</i> = <i>Ophrestia</i>	–	–	–	–
–	<i>G. max</i> (cultivated)	<i>G. max</i>	<i>G. max</i> (L.) Merr.	<i>G. max</i> (L.) Merr.
Section <i>Leptocytamus</i>	Subgenus <i>Leptocytamus</i>	Subgenus <i>Glycine</i>	Subgenus <i>Glycine</i>	Subgenus <i>Glycine</i>
<i>G. falcata</i>	<i>G. falcata</i>	<i>G. falcata</i>	<i>G. falcata</i> Benth.	<i>G. falcata</i> Benth.
<i>G. clandestina</i>	<i>G. clandestina</i>	<i>G. clandestina</i>	<i>G. clandestina</i> Wendl.	<i>G. clandestina</i> Wendl.
<i>G. clandestina</i> var. <i>sericea</i>	<i>G. clandestina</i> var. <i>sericea</i>	<i>G. clandestina</i> var. <i>sericea</i>	<i>G. clandestina</i> var. <i>sericea</i> Benth.	–
<i>G. latrobeana</i>	<i>G. latrobeana</i>	<i>G. latrobeana</i>	<i>G. latrobeana</i> (Meissn.) Benth.	<i>G. latrobeana</i> (Meissn.) Benth.
<i>G. tabacina</i>	<i>G. tabacina</i>	<i>G. tabacina</i>	<i>G. tabacina</i> (Labill.) Benth.	<i>G. tabacina</i> (Labill.) Benth.

(continued)

Table 1.1 (continued)

Bentham (1864, 1865)	Hermann (1962)	Verdcourt (1966, 1970)	Newell And Hymowttz (1980)	Current
<i>G. tabacina</i> var. <i>latifolia</i>	–	–	<i>G. latifolia</i> (Benth.) Newell and Hymowitz Hymowitz	<i>G. latifolia</i> (Benth.)
<i>G. tabacina</i> var. <i>uncinata</i>	–	–	–	–
<i>G. sericea</i>	<i>G. canescens</i>	<i>G. canescens</i>	<i>G. canescens</i> F. J. Herin	<i>G. canescens</i> F. J. Herin
<i>G. tomentosa</i>	<i>G. tomentella</i>	<i>G. tomentella</i>	<i>G. tomentella</i> Hayata	<i>G. tomentella</i> Hayata
–	–	–	–	<i>G. argyrea</i> by Tindale (1984) + 18 species (see Table 1.8)

Classified genus *Glycine* into three subgenera:

Glycine – This included the species from Africa and southeastern Asia

Soja – This included the soybean and its annual wild progenitor

Leptocyamus – Composed of nine perennial Australian species (Hermann 1962)

In 1966, Verdcourt readjusted the nomenclature and renamed *Glycine* to *Bracteata*, also subgenus *Leptocyamus* to *Glycine* (Verdcourt 1966; Hymowitz and Newell 1981). But in 1977, Lackey removed *Bracteata* as the characteristics of *G. wightii* were different from *Glycine* and hence transferred to a new genus (Lackey 1977a, b, c). Shortly after that, Newell and Hymowttz divide the genus into two subgenera (*Soja* and *Glycine*) (Newell and Hymowttz 1980). Later, 19 species were added to subgenus *Glycine* (Tables 1.1 and 1.9), while, subgenus *Soja* remains the same (Sherman-Broyles et al. 2014). More information is available at the websites of US Department of Agriculture and the Plant List:

The plant list: <http://www.theplantlist.org/tpl1.1/search?q=glycine>

US Department of Agriculture (USDA), Natural Resources Conservation Service:
<https://plants.usda.gov/java/nameSearch>

1.2.1 Soybean Morphology

The widely cultivated soybean, *G. max*, is a derivative of either *G. ussuriensis* (*soja*) or some Asiatic ancestor. It differs from other wild species in terms of inflorescence, seed pods, and stem structure. Unlike other species inflorescence has axillary

clusters or greatly reduced racemes without a bract at the base, and seed pods are broad, often curved, and contain two to four seeds. Additionally, *G. max* has a stout primary stem and sparse branches and does not twine or climb like other species (Hermann 1962). The characteristics of soybean are highly influenced by climatic conditions, soil quality, and genetics (Ratnaparkhe et al. 2011).

1.2.1.1 The Plant

Soybean is an annual plant (Ratnaparkhe et al. 2011), a bushy herb that can grow up to 2 m (Joy et al. 1998).

1.2.1.2 Roots

Soybean has a characteristic taproot system (see Fig. 1.2a). Initially, the root grows as a taproot which later has secondary, tertiary, and higher-order roots. The root nodules found in the soybean roots form a symbiotic relation with *Bradyrhizobium japonicum* (a nitrogen-fixing bacterium) and fix nitrogen in the soil (Carlson and Lersten 1987; Miladinović and Đorđević 2011), and this feature causes the legumes to have an increase in protein content (Singh and Chung 2016).

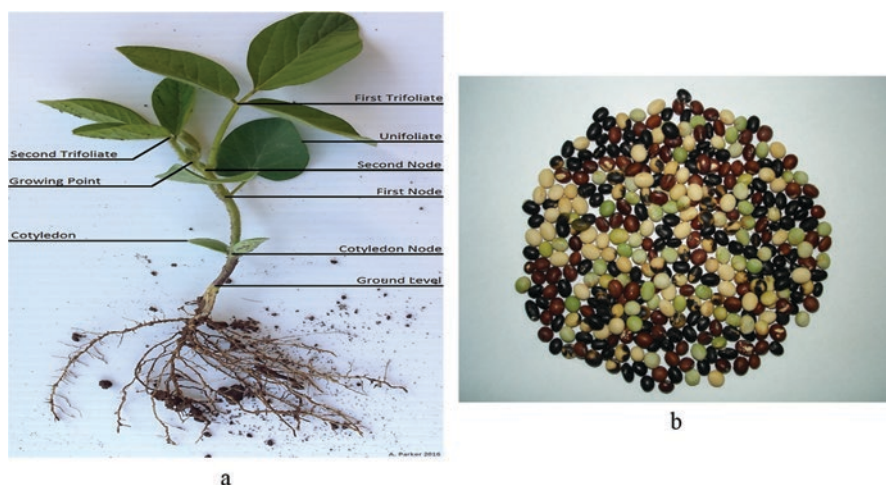


Fig. 1.2 The soybean plant. (a) Figure shows the shoot and root systems of soybean plant; photo by Adam Parker. (b) The color and shape differentiate between soybean seeds. (Photo by J. Miladinović (Miladinović and Đorđević 2011))

1.2.1.3 Stem

The stem is erect, stout, green, and covered with thick, fine brown, or gray hairs. It is mostly branched with the first node carrying the cotyledons (Fig. 1.2a) and the second node holding unifoliolate leaves. Hereinafter nodes have alternated trifoliolate leaves (Miladinović and Đorđević 2011).

1.2.1.4 Leaves

Soybean has three types of leaves – cotyledons, unifoliolate leaflets, and trifoliolates (see Fig. 1.2a). The cotyledons leaves are parts of the embryo, green or yellow, and round in shape. Unifoliolate leaves are the first pair of oval-shaped leaves. The trifoliolate leaves have three to four oval or spear-shaped leaflets per leaf. It is usually light to dark green and alternately arranged. They fall before the seeds are mature (Hermann 1962; Joy et al. 1998; Miladinović and Đorđević 2011).

1.2.1.5 Flowers

Flowers of soybean are very small (4–6 mm) that grow as a cluster of three to five flowers in the axil of a branch. The color varies from white to purple. Flowers are self-fertilizing (Miladinović and Đorđević 2011). It is a typical papilionaceous flower, with characteristic corolla composed of a standard, two wings, and two keel petals.

1.2.1.6 Fruit

The size and shape of the seed pods differ a lot with the varieties. Typically, the pods (legumes) are oblong, slightly bent (Frank and Fehr 1981) at the end, hairy, and yellowish-brown in color (Hermann 1962) and grow in clusters of 3–5 (Joy et al. 1998). Each legume contains usually one to three seeds, except that plant with *na* allele has pods with four seeds (Ratnaparkhe et al. 2011).

1.2.1.7 Seeds

Soybean seeds have different shapes (ovoid, spherical, or irregular rhomboid) and colors (Fig. 1.2b). Generally, seeds are 6–11 × 5–8 mm in diameter consisting of an embryo and a seed coat (Hermann 1962; Joy et al. 1998). The embryo is the main source of oil and protein (Wang et al. 2019a). The embryo consists of two large cotyledons, plumules, leaf primordia, epicotyl, hypocotyl, and radical.

Table 1.2 Examples of some phenotypes and their corresponding genes

Trait	Locus	Allele type	Phenotype	References
Nodulation	<i>Rj2</i>	<i>rj2(Rfg1)</i>	Causes limitation to some strains of <i>S. Fredii</i> , but does not limit <i>B. japonicum</i>	Xie et al. (2019)
Flower color	<i>W1</i>	<i>w1</i>	White flower	
Seed coat color	<i>I</i>	<i>i</i>	Colorless	
Seed coat color	<i>G</i>	<i>g</i>	Green seed color changes after maturation	
Related to soybean flowering time	<i>Gm11_10950924</i>	<i>Glyma.11G142900</i>	Transcription factor MYB59 related	Li et al. (2019)
Flowering time	<i>E1</i>	<i>E1/E1</i>	Late maturity	Owen (1927)
Stem growth habit	<i>Dt1</i>	<i>dt1</i>	Condition the determinate habit	Liu et al. (2010)

The seed coat protects the embryo. It is smooth/wrinkle in texture, glossy/matte, and water-resistant. It can be green, brown, black, or yellow. There is a scar marking on the seed coat known as hilum which is the point of attachment of funicle. The hilum is either linear or oval-shaped and can be black, brown, gray, yellow, green, or the same color as the seed coat. Hilum has a small pore at the end; micropyle, through which embryo absorbs water, exchanges gases and also serves as an outlet for radical to come out (Dzikowski 1936; Joy et al. 1998; Miladinović and Đorđević 2011).

Zhang et al. (2018) found a relationship between the shiny seed coat and the yield of oil. The seed coat of several legumes is covered with powdery bloom to protect it from predators. The seed coat bloom in wild soybean (*G. soja*) is controlled by Bloom1 (*B1*). Furthermore, Zhang et al. added that during the domestication of soybean, a nucleotide mutation in the coding region of *B1* leads to a shiny seed coat and increased the seed oil content (Table 1.2).

1.3 Soybean Cytogenetics

The soybean (*G. max*) genome is the product of a diploid ancestor ($n = 11$), which went through aneuploid loss ($n = 10$), followed by polyploidization ($2n = 20$) and diploidization ($n = 20$). It is a partially diploidized tetraploid (Lackey 1980; Singh and Hymowitz 1988). There is possibility of genome duplications or hybridization (Shoemaker et al. 1996, 2002; Blanc and Wolfe 2004; Tian et al. 2004). It is difficult to count the mitotic and meiotic chromosomes due to their smaller size. By using a chromosome image analyzing system (CHIAS), Yanagisawa et al. (1991) were able to separate 40 soybean chromosomes into 5 groups (A, B, C, D, E). Group A consists of a pair of nucleolus organizer chromosomes, group B includes two