Hongmei Miao · Haiyang Zhang · Chittaranjan Kole *Editors*

The Sesame Genome



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

Hongmei Miao • Haiyang Zhang • Chittaranjan Kole Editors

The Sesame Genome



Editors
Hongmei Miao
Henan Academy of Agricultural Sciences
Henan Sesame Research Center
Zhengzhou, China

Haiyang Zhang Henan Academy of Agricultural Sciences Henan Sesame Research Center Zhengzhou, China

Chittaranjan Kole ICAR-National Institute for Plant Biotechnology New Delhi, India

ISSN 2199-4781 ISSN 2199-479X (electronic) Compendium of Plant Genomes ISBN 978-3-319-98097-3 ISBN 978-3-319-98098-0 (eBook) https://doi.org/10.1007/978-3-319-98098-0

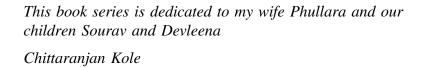
© Springer Nature Switzerland AG 2021

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

This Springer imprint is published by the registered company Springer Nature Switzerland AG The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Preface to the Series

Genome sequencing has emerged as the leading discipline in the plant sciences coinciding with the start of the new century. For much of the twentieth century, plant geneticists were only successful in delineating putative chromosomal location, function, and changes in genes indirectly through the use of a number of "markers" physically linked to them. These included visible or morphological, cytological, protein, and molecular or DNA markers. Among them, the first DNA marker, the RFLPs, introduced a revolutionary change in plant genetics and breeding in the mid-1980s, mainly because of their infinite number and thus potential to cover maximum chromosomal regions, phenotypic neutrality, absence of epistasis, and codominant nature. An array of other hybridization-based markers, PCR-based markers, and markers based on both facilitated construction of genetic linkage maps, mapping of genes controlling simply inherited traits, and even gene clusters (QTLs) controlling polygenic traits in a large number of model and crop plants. During this period, a number of new mapping populations beyond F₂ were utilized and a number of computer programs were developed for map construction, mapping of genes, and for mapping of polygenic clusters or QTLs. Molecular markers were also used in the studies of evolution and phylogenetic relationship, genetic diversity, DNA fingerprinting, and map-based cloning. Markers tightly linked to the genes were used in crop improvement employing the so-called marker-assisted selection. These strategies of molecular genetic mapping and molecular breeding made a spectacular impact during the last one and a half decades of the twentieth century. But still they remained "indirect" approaches for elucidation and utilization of plant genomes since much of the chromosomes remained unknown and the complete chemical depiction of them was yet to be unraveled.

Physical mapping of genomes was the obvious consequence that facilitated the development of the "genomic resources" including BAC and YAC libraries to develop physical maps in some plant genomes. Subsequently, integrated genetic—physical maps were also developed in many plants. This led to the concept of structural genomics. Later on, emphasis was laid on EST and transcriptome analysis to decipher the function of the active gene sequences leading to another concept defined as functional genomics. The advent of techniques of bacteriophage gene and DNA sequencing in the 1970s was extended to facilitate sequencing of these genomic resources in the last decade of the twentieth century.

viii Preface to the Series

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 travelled much beyond biology and involves biophysics, biochemistry, and bioinformatics!

Thanks to the efforts of both public and private agencies, genome sequencing strategies are evolving very fast, leading to cheaper, quicker, and automated techniques right from clone-by-clone and whole-genome shotgun approaches to a succession of second-generation sequencing methods. The development of software of different generations facilitated this genome sequencing. At the same time, newer concepts and strategies were emerging to handle sequencing of the complex genomes, particularly the polyploids.

It became a reality to chemically—and so directly—define plant genomes, popularly called whole-genome sequencing or simply genome sequencing.

The history of plant genome sequencing will always cite the sequencing of the genome of the model plant Arabidopsis thaliana in 2000 that was followed by sequencing the genome of the crop and model plant rice in 2002. Since then, the number of sequenced genomes of higher plants has been increasing exponentially, mainly due to the development of cheaper and quicker genomic techniques and, most importantly, the development of collaborative platforms such as national and international consortia involving partners from public and/or private agencies.

As I write this preface for the first volume of the new series "Compendium of Plant Genomes," a net search tells me that complete or nearly complete whole-genome sequencing of 45 crop plants, eight crop and model plants, eight model plants, 15 crop progenitors and relatives, and three basal plants is accomplished, the majority of which are in the public domain. This means that we nowadays know many of our model and crop plants chemically, i.e., directly, and we may depict them and utilize them precisely better than ever. Genome sequencing has covered all groups of crop plants. Hence, information on the precise depiction of plant genomes and the scope of their utilization are growing rapidly every day. However, the information is scattered in research articles and review papers in journals and dedicated Web pages of the consortia and databases. There is no compilation of plant genomes and the opportunity of using the information in sequence-assisted breeding or further genomic studies. This is the underlying rationale for starting this book series, with each volume dedicated to a particular plant.

Plant genome science has emerged as an important subject in academia, and the present compendium of plant genomes will be highly useful to both students and teaching faculties. Most importantly, research scientists involved in genomics research will have access to systematic deliberations on the plant genomes of their interest. Elucidation of plant genomes is of interest not only for the geneticists and breeders, but also for practitioners of an array of plant science disciplines, such as taxonomy, evolution, cytology,

Preface to the Series ix

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

Sesame (Sesamum indicum L., 2n = 26) is an ancient and widely grown oilseed crop with high oil and antioxidants content. In China, sesame has been used as medicine and was mentioned in the earliest Chinese medicine dictionary 'ShenNong's herb classic' more than 2000 years ago. In India and other ancient countries, sesame has been applied for food, medicine, and religious ceremonies. The vigor of the endless flowering of sesame is always applied to express the bright expectation of one's life by the Chinese people. The story 'Ali Baba and the Forty Thieves' and the famous magic code 'Open Sesame' also suggest that sesame was very important for the progress of civilization, especially in the ancient Arabic regions. At present, sesame is widely cultivated in 80 countries distributed mainly in the tropical and subtropical regions of the world for its high tolerance to high temperature and arid environments. However, sesame is a traditional crop and the sole cultivated species in the genus Seamum of the Pedaliaceae family. The relative narrow genetic base in sesame limits the development of the genetics and breeding research on it. Prof. Amram Ashri has considered sesame as an orphan species for the remote phylogenetic relationship with other crops and the wild Sesamum species and the rare knowledge on its genetics.

The initiation of the Sesame Genome Project by the Sesame Genome Working Group in 2010 was the prelude to sesame genomics and comparative genomics research. The achievements of the sesame genomics and functional genomics research provided the necessary genome information and improved the sesame genetics and breeding studies around the world. Thus, we summarize the main progresses in sesame genome research, as well as the necessary background knowledge of sesame genetics and genomics research to readers in this book entitled, *The Sesame Genome*.

This book comprises 19 chapters, which present the world sesame production status, description of botanical traits, classical genetics and molecular genetics of key agronomic traits, cytological and chromosome group analysis, techniques of mutation and creation of genetic variation, breeding techniques and popular improved varieties, and the achievements of genome sequencing and genomics research in sesame. We also introduce the background of initiation of Sesame Genome Project, the strategies and tools of sesame genome sequencing and assembly, genome structure, key gene families, and genome evolution of sesame. The genome sequencing of the wild *Sesamum* species is also enumerated in the book for the first time.

xii Preface

We sincerely attempted to edit the book as a useful tool and guide for sesame scientists and industries. In order to increase the visuality and instructiveness of the contents, a great number of unpublished figures and photographs are included by authors. Lots of new and unpublished research results are also provided. Moreover, the new objectives of sesame breeding and the concepts of the ideal sesame put forward by Prof. Haiyang Zhang, Mr. D. Ray Langham, and other famous sesame scientists are listed and discussed in the book. With the continuous development of the genome sequencing techniques, the quality of the sesame genome assembly will obviously be improved in near future. More tools and data resources will be established accordingly. Therefore, we hope that the scientific information and experimental ideas reflected in this book are helpful for future molecular genetics and genomics research in sesame and other crops.

This book is the first professional work about sesame genome. The contents also exhibit the fruits of the international cooperation and communication on sesame genetics, genomics and breeding research in recent years. We expect that the application of the updated genome and genetics information would improve the collaboration of the world sesame research and enhance the development of the world sesame industry in the future.

Zhengzhou, China Zhengzhou, China New Delhi, India Hongmei Miao Haiyang Zhang Chittaranjan Kole

Acknowledgments This work was financially supported by the earmarked fund for China Agriculture Research System (CARS-14), the Key Project of Science and Technology of Henan Province (201300110600); Zhongyuan Science and Technology Innovation Leading Talent Plan (214200510020), the Henan Province Specific Professor Position Program (SPPP2020), the Zhongyuan Scientist Workshop Construction (ZSWC2019; 214400510026), and the Innovation Scientists and Technicians Troop Construction Projects of Henan Province (ISTTCPHP2016).

Contents

1	Haiyang Zhang, D. Ray Langham, and Hongmei Miao	1
2	Botanical Descriptions of Sesame	19
3	Beneficial Components in Sesame Proteins and Oil Jason T. C. Tzen	59
4	Classical Genetics of Sesame. Libin Wei, Hongmei Miao, Yinghui Duan, and Haiyang Zhang	7 9
5	Mutagenesis for Creation of Genetic Variability in Sesame	121
6	Tissue Culture and Genetic Transformation in Sesame Hongmei Miao, Ming Ju, Huili Wang, and Haiyang Zhang	131
7	Traditional Breeding in Sesame	145
8	Molecular Mapping and Breeding in Sesame Haiyang Zhang, Yingzhong Zhao, Libin Wei, Ting Zhou, Yinghui Duan, and Hongmei Miao	159
9	Cytological Details of Sesame	179
10	Background of the Sesame Genome Project Haiyang Zhang, Lei Wang, and Hongmei Miao	199
11	Sequencing of Chloroplast Genome in Sesame Chun Li, Haiyang Zhang, Ming Ju, Fangfang Xu, and Hongmei Miao	205
12	Strategies and Tools for Sequencing of the Sesame Genome Hongmei Miao, Yamin Sun, Lei Wang, and Haiyang Zhang	209
13	Sesame Genome Assembly	225

xiv Contents

14	Repetitive Sequences in Sesame Genome	239
15	Genome Annotation and Gene Families in Sesame	255
16	Genome Synteny Analysis and Phylogenetic Position of <i>S. indicum</i> Hongmei Miao, Yamin Sun, Wenchao Lin, Lei Wang, and Haiyang Zhang	267
17	Genome Sequencing of the Wild Sesamum Speices	275
18	The Sesame Genome for Gene Discovery in Sesame Yinghui Duan, Ming Ju, Hongmei Miao, and Haiyang Zhang	283
19	Prospect of Designed Breeding in Sesame in the Post-genomics Era. Haiyang Zhang, D. Ray Langham, and Hongmei Miao	291

Contributors

Yinghui Duan Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, People's Republic of China

Ming Ju Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, People's Republic of China

Ali Khalafalla Agriculture Research Corporation, Wad Madani, Sudan

D. Ray Langham Sesame Research, LLC, San Antonio, USA

Chun Li Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, People's Republic of China

Wenchao Lin TEDA School of Biological Sciences and Biotechnology, Nankai University, Tianjin, People's Republic of China

Qin Ma Henan Academy of Agricultural Sciences, Henan Sesame Research Center, Zhengzhou, People's Republic of China

Hongmei Miao Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, People's Republic of China

Yamin Sun TEDA School of Biological Sciences and Biotechnology, Nankai University, Tianjin, People's Republic of China

Jason T. C. Tzen Graduate Institute of Biotechnology, National Chung-Hsing University, Taichung, Taiwan

Huili Wang Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, People's Republic of China

Lei Wang TEDA School of Biological Sciences and Biotechnology, Nankai University, Tianjin, People's Republic of China

Libin Wei Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, People's Republic of China

Fangfang Xu Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, People's Republic of China

Haiyang Zhang Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, People's Republic of China

xvi Contributors

Ruihong Zhao Henan Academy of Agricultural Sciences, Henan Sesame Research Center, Zhengzhou, People's Republic of China

Yingzhong Zhao Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan, People's Republic of China

Ting Zhou Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan, People's Republic of China

Abbreviations

AAR Average Arm Ratio

AFLP Amplified Fragment Length Polymorphism

AKC Asymmetric Karyotype Coefficients
ARC Agricultural Research Corporation
BAC Bacterial Artificial Chromosome

BN Branch Number

BSA Bulked Segregant Analysis

BSG Barium Hydroxide, Salt and Giemsa

BUSCO Benchmarking Universal Single-Copy Orthologs

CA Celera Assembler

CEGMA Core Eukaryotic Genes Mapping Approach

CLR Continuous Long Reads
CN Capsule Number per Plant
CRP Coordinated Research Project
CSPGs Chondroitin Sulfate Proteoglycans

CSR Charcoal Stem Rot
DAF Days After Flowering
DAP Days After Planting
DDBJ DNA Data Bank of Japan

EIAR Ethiopian Institute of Agricultural Research

EMS Ethylmethane Sulfonate
ENA European Nucleotide Archive
ESTs Expressed Sequence Tags

EST-SSR Expressed Sequence Tag-Simple Sequence Repeat

EVM Evidence Modeler

FISH Fluorescence in Situ Hybridization FOS Fusarium Oxysporum f. sp. Sesami

GCA General Combining Ability
GFP Green Fluorescent Protein

GISH Genome Fluorescence in situ Hybridization

GLM General Linear Model

GMO Genetically Modified Organism

GMS Genetic Male Sterile GO Gene Ontology

GWAS Genome-wide Association Analysis

HFC Height to the First Capsule

HGAP Hierarchical Genome Assembly Process

xviii Abbreviations

HGP Human Genome Project

Hi-C High-Throughput Chromosome Conformation Capture

HSRC Henan Sesame Research Center
IAR The Institute of Agricultural Research

ICBN International Code of Botanical Nomenclature

IDH Isocitrate Dehydrogenase

IDRC International Development Research Centre

InDel Insertion/Deletion

IPGRI International Plant Genetic Resources Institute

ISSR Inter-Simple Sequence Repeat

KEGG Kyoto Encyclopedia of Genes and Genomes

KIN Kinase

LD Linkage Disequilibrium

LG Linkage Group

LINE Long Interspersed Nuclear Element

LN Leaf Number per Plant LPS Lipopolysaccharide LTR Long Terminal Repeat

LTR-RT Long Terminal Repeat-Retrotransposon

MAS Marker-Assisted Selection

MCIM Mixed Linear Composite Interval Mapping

MCM1 Mini-Chromosome Maintenance 1

MIM Multiple Interval Mapping

MIPS Myo-Inositol 1-Phosphate Synthase

NBPGR National Bureau of Plant Genetic Resources
NBS-LRR Nucleotide Binding Site-Leucin Rich Repeat
NCBI National Center for Biotechnology Information

ncRNAs Noncoding RNAs ND Nondehiscent

NGS Next-generation Sequencing
NMR Nuclear Magnetic Resonance
NPTII Neomycin Phosphotransferase II
OLC Overlap Layout Consensus
OLP Osmotin-like Protein

PCR Polymerase Chain Reaction

PEG Polyethylene Glycol

PH Plant Height

PRG Plant Resistance Gene
PUFA Polyunsaturated Fatty Acids
QTL Quantitative Trait Locus

RAD-seq Restriction-site Associated DNA Sequencing RAPD Random Amplified Polymorphic DNA

rDNA Ribosomal DNA

RIL Recombinant Inbred Line RLK Receptor-Like Kinase

RLP Resistance to powdery mildew

RRGS Reduced Representation Genome Sequencing

Abbreviations xix

RSAMPLs Random Selective Amplification of Microsatellite

Polymorphic Loci

RT-PCR Reverse Transcription—Polymerase Chain Reaction

SCA Specific Combining Ability

SCAR Sequence Characterized Amplified Region

SEM Scanning Electron Microscope

SFW Sesame Fusarium Wilt SGP Sesame Genome Project

SGWG Sesame Genome Working Group

SGWG-BD The Sesame Genome Working Group of Bioengineering

Department

SGWG-PPRI The Sesame Genome Working Group of Plant Protection

Research Institute

SINE Short Interspersed Nuclear Element

SLAF-Seq Specific Length Amplified Fragment Sequencing

SMRT Single Molecule Real-Time
SN Seed Number per Capsule

SNP Single Nucleotide Polymorphism

SRAP Sequence-Related Amplified Polymorphism

SRP Signal Recognition Particle

SSC Small Single-Copy

SSR Simple Sequences Repeats
TAMRA Tetramethyl-Rhodamine-5-dUTP
TBC Tianjin Biochip Corporation
TEM Transmission Electron Microscopy

TEs Transposable Elements
TF Transcription Factor
TNL Tamil Nadu Local

TSBSB TEDA School of Biological Sciences and Biotechnology

TSW Thousand Seed Weight UE Upper Epidermis

UPGMA Unweighted Pair-Group Method with Arithmetic Averages

WGS Whole Genome Sequencing

Economic and Academic Importance of Sesame

Haiyang Zhang, D. Ray Langham, and Hongmei Miao

Abstract

Sesame is an ancient oilseed crop. In this chapter, we introduce the status of sesame production and the world trade, cultivation styles, and seed nutrition. With the increase of harvested area and the yield level, the world sesame production is accelerating. The increasing demand in the market is stimulating the development of the sesame industry. In sesame seeds, high unsaturated fatty acids, proteins, minerals, and antioxidants result in the high value of sesame products. The nutritional and pharmaceutical advantages of sesame in food industry and medicine industries are concisely recognized and valued.

H. Zhang (☒) · H. Miao Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, China e-mail: zhanghaiyang@zzu.edu.cn

D. R. Langham

Sesame Research, LLC, San Antonio, USA e-mail: raylangham@sesameresearch.org

H. Zhang · H. Miao Henan Provincial Key Laboratory of Specific Oilseed Crops Genomics, Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, China

1.1 Introduction

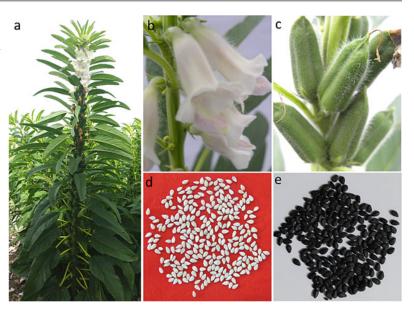
(Sesamum indicum L., 2n = 26(Fig. 1.1) belongs to the Sesamum genus of the Pedaliaceae family and is one of the five main oilseed crops of the world. Sesame seeds are traditionally used for oil crushing and food consumption for the high oil (29.5-62.7%) and protein (12.9-30%) contents. S. indicum is the sole cultivar in the genus Sesamum (Fig. 1.1). The cultivation history can be traced back to 3050-3500 BC in the Harappa Valley of the Indian subcontinent (Bedigian and Harlan 1986). Sesame is tolerant to drought and arid environments. At present, sesame is mainly cultivated in the tropical and subtropical regions of Asia, Africa, and America continents (Ashri 1998).

1.2 World Sesame Production

Sesame is a traditional oilseed crop and is mainly cultivated in more than 80 countries of the world (Fig. 1.2). About 70% of the regions for sesame production located in the latitudes from 30° S to 43° N. All these countries are primarily located in Africa, Asia, Central America, and Latin America under hot and dry environments. Especially, Africa is an important region for sesame production. Even though the progenitor of sesame is not determined and the origins of sesame are still debated (Bedigian et al. 1985;

[©] Springer Nature Switzerland AG 2021 H. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,

Fig. 1.1 Images of sesame plant, flower, capsule, and seeds. a Sesame (Chinese var. Yuzhi Dw607 with short internode length) plant in field. b Sesame flowers. c Sesame capsules. d and e White- and black-coated sesame seeds. (Photographs provided by Haiyang Zhang)



Kobayashi 1986; Ashri 1998), Africa has most of the wild *Seamum* species and represents the traditional cultivation history of sesame of the world.

At present, a dozen of countries in the eastern or western Africa such as Sudan, Nigeria, Tanzania, Ethiopia, Burkina Faso, South Sudan, Chad, and Uganda produce more than 100 Kilo tons sesame seeds per country annually (2018 FAO data, www.fao.org/statistics/en/). In Asia,

the main production countries include India, Myanmar, and China. Mexico, Paraguay, Bolivia, Venezuela, Guatemala, Honduras, Nicaragua, and the United States in America supply a smaller portion of sesame production, but these countries, particularly the USA, Venezuela, Paraguay, and Bolivia, are the leaders in mechanized harvest technology. Moreover, Portugal, Greece, Italy, and Bulgaria in Europe also plant sesame for domestic consumption in European

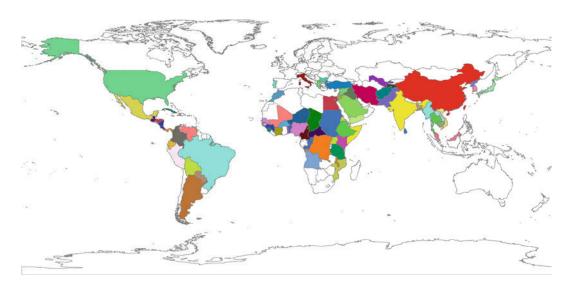


Fig. 1.2 Distribution of the world sesame production. Eighty sesame production countries are shown in color

Table 1.1 Sesame production statistics of the world and main production countries in 2018

Country name	Area harvested(Ha)	Production(Tons)	Yield(Kg/ha)
World (Total)	11,743,382	6,015,573	5123
Sudan	3,480,000	981,000	2819
India	1,730,000	746,000	4312
Myanmar	1,463,447	768,858	5254
Tanzania	800,000	561,103	7014
South Sudan	617,908	206,522	3342
Nigeria	539,089	572,761	10,625
Burkina Faso	438,941	253,936	5785
Ethiopia	415,112	301,302	7258
Chad	325,114	172,539	5307
China	311,203	433,386	13,926
Others	1,622,568	1,018,166	6275

All the above data cited from FAO dataset (www.fao.org/statistics/en/)

Union countries in recent years (Ray Langham, personal communication).

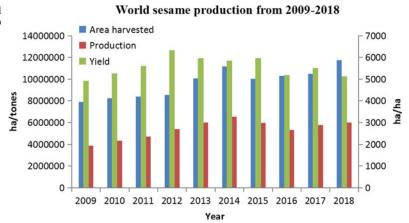
In 2018, the total area of sesame harvested in the world was 11.74 million hectares, and annual production was 6.02 million tons (Table 1.1) (FAO data). The top five countries with high harvest area for sesame include Sudan, India, Myanmar, Tanzania, and South Sudan. China had the highest sesame yield of 1392.6 kg/ha in 2018. However, the average yield of the world sesame is still low at 512.3 kg/ha.

Currently, all the top ten countries for sesame production are developing countries in the Asia and Africa continents (Table 1.1). Some developed countries, such the USA and Portugal, also grow sesame in arid regions for domestic

consumption. In USA, there are about 60,000 hectares with an increase trend in recent few years (data supplied by D. Ray Langham).

With the rapid development of the world agriculture, sesame production has increased in the past decade (Fig. 1.3). The world harvested sesame area rose from 7.87 million hectares in 2009 to 11.74 million hectares in 2018, with the mean annual increase of 4.25% (ranging from 0.28 to 14.94% from 2009 to 2018). The world sesame production increased from 3.87 million tons in 2009 to 6.02 million tons in 2018, with the mean annual increase of 4.44% (ranging from 12.23 to 12.84%). Meanwhile, the average yield of sesame annually increases 2.77% during 2009–2018. The data suggest that the increase of

Fig. 1.3 Trends in the world sesame production from 2009 to 2018



world sesame production is mainly resulted from the increase of world harvest area (Zhang et al. 2019). Africa is the sole region with an increase tendency for sesame production.

1.3 World Sesame Trade

The world sesame trade market has been stimulated by the increased sesame production in the recent years. From 2008 to 2017, the total trade amount of the main export countries ranged from 0.9575 million tons (2008) to 1.8617 million tons (2017). The imported amount of sesame seeds ranged from 1.081 million tons (2008) to 1.9085 million tons (2017) (Fig. 1.4). It is worth noting that China is the largest sesame importer since 2003 for the decrease of total production and harvest area. In the past ten years, the annual imported sesame product touched to 0.977 million tons (2016) (www.fao.org/statistics/en/). Even though the domestic sesame seeds contain high oil content with high appearance quality and good flavor, the sesame harvested area in China gradually declined for the low profit and high labor cost of sesame production. Moreover, massive labor transfer from the country to cities and low trade price in the world sesame market enhance the decrease of sesame production in China.

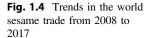
Globally, the trade prices fluctuated in the past decade (Fig. 1.5). The import price of sesame

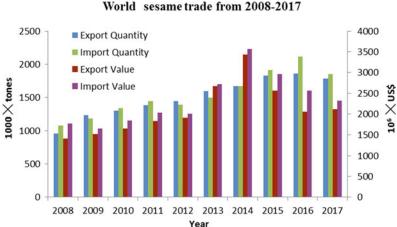
seeds in China varied from \$714 per ton in 2003 to \$1027 per ton in 2016. The highest price of \$1751 per ton was in 2014. The import price in Japan followed the same pattern. Due to unexpected weather and total annual production variation in the main sesame producing regions, the international trade market presents the instability and always influences the import price. Subsequently, the great variation of price influences the enthusiasm of the sesame farmers, even though the sesame processors (both for oil and edible products) have a small profit margin from the sesame trade fluctuation. Finally, this instability affects the development of the global sesame industry.

1.4 Sesame Cultivation

Sesame is a survivor crop, because sesame is mainly cultivated under low input conditions in most of the main production countries (Langham 2007). However, the planting regions for sesame cover different regions from irrigated to rain-fed conditions. In different production regions, the models and techniques for sesame cultivation vary accordingly. The main planting models include one cropping, double cropping, triple cropping per year, intercropping, thin population, and dense population (Fig. 1.6).

There are two types of sesame in the world according to the levels of moisture and fertility





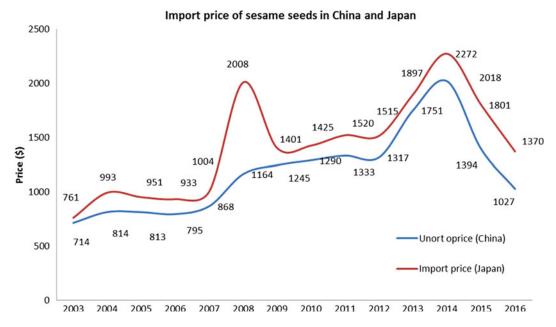


Fig. 1.5 Import price variation of sesame seeds in China and Japan from 2003 to 2016. Source from Chinese customs dataset



Fig 1.6 Representative planting models of sesame in China. **a** One cropping. **b** Double cropping. **c** Triple cropping. **d** Intercropping. (Photographs provided by Haiyang Zhang and Hongmei Miao)

(Langham 2007). One is cultivated in dry areas or seasons (such as in USA, Venezuela, Ethiopia, and Sudan) (Fig. 1.7c, d). The other is mainly cultivated under conditions with high moisture and fertilizer input (such as in China and Korea) (Fig. 1.7a, b). Meanwhile, one variety can present different architecture, when planted under high and low input architectures, respectively. In the high input architectures using high moisture and fertility, the plants grow strong during the vegetative and reproductive phases. The leaves are large; the internodes are long; the height to the first capsule is high; and there are more capsule node pairs. On the contrary, the low input architecture with low moisture and/or fertility results in the small plant type. The leaves are small; the internodes are short; the height of the first capsule is low; and there are fewer capsule node pairs. Furthermore, the capsule number per plant and the seed number per capsules are low. The yield is normally lower than that of the high input conditions. The ability of some sesame genotypes to withstand drought allows it to be one of the few crops that can be cultivated in extreme dry areas of the world.

Sesame is an indeterminate crop with long flowering time. Langham (2007) divided the phenology of sesame into four main phases

(vegetative, reproductive, ripening, and drying) and ten stages (Table 1.2). Reproductive, ripening, and drying phases generally overlap. For example, sesame plants can still be flowering, while the lower capsules are ripe and shattering. In the main production regions, all the farming actions (planting, thinning, weeding, cutting, shocking, and drying) are still manual. Thus, the drying phase is meaningful only for the plants that are left in the field to dry for direct harvest with a combine.

The greatest advances in modern agriculture lie in the mechanized practices from sowing until the harvest phase. Sesame is traditionally a manual crop. This laborious planting model and low profit are not suitable for the rapid development of the world agriculture. Stabilizing the high yield and increasing the relative profits are the key and everlasting objective of sesame production. Raising the mechanization level is necessary for improving sesame production. In 1943 in Venezuela, D.G. Langham introduced mechanization with mechanical planting, cultivating, cutting and binding, and combining. The only manual operation was in shocking the sesame and feeding the shocks into the combine. In 1982 in the USA, Sesaco Company released varieties that could be swathed into a windrow and left in the

Fig. 1.7 Representative sesame varieties planted in various conditions. a Sesame variety Yuzhill in China. b Sesame variety in South Korea. c Sesaco sesame variety in USA. d Sesame variety in Sudan (Photographs provided by Haiyang Zhang, Churl Whan Kang, Ray Langham, and Khalafalla Ali, respectively)



Table 1.2 Phases and stages of sesame

Stage/phase	ge/phase End point of stage		Week number
Vegetative			
Germination	Emergence	0–5	1-
Seedling	Third pair true leaf length = 2nd	6–25	3-
Juvenile	First buds	26–37	2-
Pre- reproductive	50% open flowers	38–44	1-
Reproductive			
Early bloom	Five node pairs of capsules	45–52	1
Mid bloom	bloom Branches/minor plants stop flowering		4
Late bloom	90% of plants with no open flowers	82–90	1+
Ripening	Physiological maturity (PM)	91–106	2+
Drying			
Full maturity		107– 112	1-
Initial drydown	drydown First dry capsule		2
Late drydown Full drydown		137– 146	3

Source from Langham (2007)

field to dry and then harvested by a combine. This was the first sesame that was completely mechanized. In 1988 in the USA, Sesaco released the first sesame varieties that could be harvested directly in the field after the sesame was dry enough for almost 6% seed moisture. In 1997, nondehiscent (ND) sesame variety was released in the USA with improved shatter resistance. In 2008, improved nondehiscent sesame variety was released. After being dry one month, 85% of the capsules in improved nondehiscent plants still have seed to the top (Langham 2008). At present, the above new sesame varieties with the increased shatter resistance trait are widely cultivated with the mechanized technology in the USA, Paraguay, and Bolivia (Fig. 1.8). We believe that the mechanization innovation of sesame cultivation will improve the world sesame industry in the near future.

1.5 Sesame Nutrition

Sesame is a high oil content crop. In the principal producing countries, sesame plays a considerable role in the development of the national economy, even though sesame contributes only about 1% of the total world oilseed production. Compared with cereals and other oilseed crops, sesame seed high oil and dietary fiber contents (Table 1.3). The nuclear magnetic resonance (NMR) evaluation results indicate that the oil content in the 50 foreign and domestic sesame germplasm resources ranges from 29.48 to 58.71% (Wei et al. 2016). Uzun et al. (2008) reported specific samples with 62.7% oil content. Besides abundant protein (12.9–30%), sesame seed is also rich in vitamins and minerals (Table 1.4). Of the eight kinds of vitamins

^aDAP = days after planting. Data were collected based on S26 in 2004 at Uvalde, Texas, under irrigation condition

Fig. 1.8 Mechanized cultivation with improved nondehiscent sesame varieties in the United States.
a Improved nondehiscent sesame variety with greater retention of seeds in capsule.
b Mechanized cultivation.
c Capsule mature before drydown. d Direct combine harvest (Photographs provided by D. Ray Langham)



Table 1.3 Nutritional ingredients in sesame and other crops (g per 100 g)

Crop	Oil	Protein	Carbohydrate	Crude fiber	Moisture
Sesame (white- seeded)	52.0	18.4	21.7	9.8	5.3
Peanut seed	50.0	25.8	16.2	_	5.5
Soybean	21.0	35.0	18.7	4.8	15.5
Rapeseed	38.0	26.0	24.0	_	8.0
Wheat flour	1.8	9.9	74.6	0.6	12.0
Corn (yellow)	4.3	8.5	72.2	1.3	12.0

Data are cited from Zhang et al. (2012)

Table 1.4 Vitamin and mineral contents in sesame seeds

Vitamin component ^a	Content (mg/100 g)	Mineral component ^b	Content (mg/100 g)
Thiamine (B ₁)	1.3	Calcium (Ca ²⁺)	1345
Niacin	5.0	Magnesium (Mg ²⁺)	344
Vitamin C	0.5	Potassium (K ⁺)	674
Lactochrome (B ₂)	6.3	Iron (Fe ³⁺)	69
Pantothenic acid	600	Sodium (Na ⁺)	75
Vitamin E	50	Copper (Cu ²⁺)	19
Tocopherol (α, β)	5	Manganese (Mn ²⁺)	17
Folic acid	18.5	Selenium (Se ⁶⁺)	0.4
		Zinc (Zn ²⁺)	37

^aVitamin content data are cited from Zhang et al. (2012)

^bMineral content data are supplied by Henan Sesame Research Center, Henan Academy

existing in sesame seeds, the content of pantothenic acid (600 mg in 100 g sample) ranks no. 1, followed by vitamin E (50 mg), folic acid (18.5 mg), lactochrome (B₂) (6.3 mg), niacin (5.0 mg), tocopherol (α , β) (5 mg), thiamine (B₁) (1.3 mg), and vitamin C (0.5 mg). In addition, in 100 g sesame seeds, there are 1345 mg calcium, 344 mg magnesium, 674 mg potassium, 69 mg iron, 75 mg sodium, 37 mg zinc, 19 mg copper, 17 mg manganese, and 0.4 mg selenium. Furthermore, sesame oil contains many minerals like magnesium, copper, calcium, iron, and zinc, and vitamins like B₆. The bulk of these minerals in the sesame seed and sesame oil obtain the high acceptance among the customers (Prakash and Naik 2014). In sesame leaves, the concentration of calcium, zinc, iron, and magnesium is 19,212.6 mg/kg, 15.7 mg/kg, 301.5 mg/kg, 1433.9 mg/kg, respectively (data provided by Zhigang Liu). In China, sesame leaves are becoming a popular vegetable. In Tanzania, the leaves of a wild sesame species, namely 'Mlenda', have been used as a vegetable.

In addition, the sesame seed carbohydrate content is high to 25.0%, while the digestible fiber in sesame seeds reaches to 9.8 g per 100 g (Jimoh and Aroyehun 2011; Zhang et al. 2012; Makinde and Akinoso 2013; Prakash and Naik 2014; Sene et al. 2017). At present, sesame seeds are applied as the preferred food, especially in developed Asian countries.

1.5.1 Fatty Acid Composition in Sesame Oil

Sesame oil is composed of triglyceride and a little amount of phospholipid (0.03-0.13%). There are saturated, monosaturated, and polysaturated fatty acids in sesame oil (Zhang et al. 2019). Oleic acid (18:1) ($\sim 40\%$) and linoleic acid (18:2) ($\sim 46\%$) are the main fatty acids and consist of about 85% of the sesame oil. In addition, the saturated fatty acids include palmitic acid (16:0) (8%) and steric acid (18:0) (4%). The content of arachidic acid (20:0) and linolenic acid (18:3) is below 1%. For

polysaturated acids, linoleic acid especially is a requisite fatty acid and cannot be synthesized in human body. Linoleic acid can inhibit the synthesis of cholesterol in human blood and soften the blood vessel to prevent arteriosclerosis. In India, the sesame oil is recommended for topical use by Ayurveda, as sesame oil contains selective antineoplastic properties which are similar to those demonstrated for essential polyunsaturated fatty acids and their metabolites (Smith and Salerno 1992). Smith and Salerno (1992) proved that that linoleic acid of sesame and safflower oils could selectively inhibit malignant melanoma growth over normal melanocytes, whereas coconut, olive, and mineral oils contain little or no linoleic acid at all. Therefore, sesame is also believed as the ideal vegetable oil for body health.

1.5.2 Amino Acid Composition in Sesame Seeds

Sesame seeds contain 18–20% proteins (Zhang et al. 2019). Li et al. (2014) evaluated the 369 sesame accessions and determined that the protein content in sesame seeds varied from 16.72 to 27.79%. Insoluble 11S globulin and soluble 2S albumin are termed as α -globulin and β -globulin, respectively, and are the two major storage proteins in sesame seed. The two proteins occupy 80-90% of total seed proteins (Anilakumar et al. 2010). Comparison of amino acid compositions indicates that 11S globulin and soluble 2S are substantially less hydrophobic than the known oleosins and thus should not be aggregated multimers of oleosins. The functional properties of globulin in sesame can be influenced by interactions with food components and the processing treatments (Anilakumar et al. 2010). Sesame seed contains 18 types of amino acids (Table 1.5). In white sesame seeds, the main amino acids are glutamic acid (4.21%) and arginine (2.79%) (Zhang et al. 2019). Sesame is rich in sulfur containing amino acids and limited in tryptophan (Kapadia et al. 2002; Anilakumar et al. 2010; Gao et al. 2011; Prakash and Naik 2014).

Table 1.5 Amino acid composition in sesame seeds

Amino acid component	Content (%)		
	White seed ^a	Black seed ^a	
Glutamic acid	4.21	4.02	
Arginine	2.79	2.60	
Aspartic acid	1.66	1.62	
Leucine	1.41	1.33	
Phenylalanine	1.02	0.96	
Valine	1.01	0.97	
Glycine	1.00	0.98	
Alanine	0.90	0.88	
Гугоѕіпе	0.90	0.84	
Γhreonine	0.89	0.86	
Methionine	0.84	0.61	
soleucine	0.77	0.72	
Serine	0.73	0.71	
Lysine	0.67	0.69	
Proline	0.58	0.54	
Histidine	0.47	0.47	
Cystine	0.37	0.36	
Гryptophan	0.21	0.20	
Гotal	20.42	19.33	

^aWhite sesame seeds (var. Yuzhi 11) and black seeds (var. Ji9014) harvested in Yuanyang experimental station in 2009 are evaluated using standard chemical method by HSRC, HAAS, China (cited from Zhang et al. 2019)

1.5.3 Oxidants in Sesame Seeds

Antioxidants are molecules which present with low content but prevent or reduce the extent of oxidative destruction of biomolecules, according to the classical definition (Halliwell 1990). In the famous Mediterranean diet pyramid, olive oil is recommended, as it contains 2% unsaponifiable fraction. The most abundant antioxidants in olive oil are lipophilic and hydrophilic phenols. Sesame seed has been used as a traditional health food and medicine to prevent numerous diseases in Africa and Asia since the dawn of civilization (Shittu et al. 2007). In Chinese medicine dictionaries, such as 'ShenNong's herb classic' (the earliest Chinese medicine work, written more than 2000 years ago), 'Supplementary records of famous physician' (formed at the 2nd or 3rd century AD), and 'Wu Pumedica' (edited at the third century AD), sesame has been recorded as a tonic and medicine. The active components in sesame seeds refer to antioxidants, such as lignans (Ram et al. 1990; Baydar 2005; Suwimol et al. 2012). Sesame shows a remarkable stability to oxidation (Abou-Gharbia et al. 1997, 2000) and is regarded as 'the queen of the plant oil crop seeds', perhaps for the high resistance to oxidation and rancidity (Bedigian and Harlan 1986).

Comparison of unroasted and roasted sesame seed oil indicates that γ -tocopherol is the main active constituent in fresh unroasted seed oil, while sesamol has high concentration in roasted seed oil, which is produced by hydrolysis of sesamolin. γ -tocopherol is the major tocopherol in sesame seeds, and the content is high to be 490–680 mg/kg sesame oil. In wild sesame species of *S. alatum*, *S. angustifolium*, and *S. latifolium*, the amounts of γ -tocopherol are 210–320, 750, and 800 mg/kg sesame oil, respectively. α -tocopherol is the predominant form in

the photosynthetic tissues such as stems and leaves (Kiani et al. 2016). The main function of α -tocopherol is to be a radical-chain breaking antioxidant in membranes and lipoproteins, as well as in foods (Kamaleldin and Appelqvist 1996). α -tocopherol is more useful for decreasing platelet aggregation, low-density lipid oxidation, and delaying intra-arterial thrombus formation.

Lignans and glucosides are present in the sesame seeds. Lignans are oil soluble and mostly are in sesame oil, whereas glucosides are water soluble and mostly are in the meal. Sesame lignans are comprised of sesamin, sesamolin, and small amounts of sesaminol, piperitol, sesamolinol, pinoresinol, (+)-episesaminone, hydroxymatairesinol, allohydroxymatairesinol, larisiresinol (Shittu et al. 2007). Sesamin and sesamolin are the main lignans in sesame seeds but have no antioxidative ability. Other components such as sesaminol carry free phenolic groups and present the antioxidant activity (Kanu et al. 2010) (Fig. 1.9). Physical characters and the main functions of lignans are listed in Table 1.6.

In the 1380 Chinese domestic and world sesame germplasm accessions, the lignan content varies from 0.517 to 15.832 mg/g (Haiyang Zhang, unpublished data). The content of sesamin and sesamolin ranges from 0.200 mg/g to 10.598 mg/g and from 0.024 mg/g to 7.521 mg/g, respectively (Fig. 1.10). In sesame seeds, the average sesamin content is a little higher than that of sesamolin. Most varieties in production have about 5.0 mg/g lignans in seeds.

In addition, we compared the lignan content of the cultivated sesame (var. Yuzhi 11) and five wild Sesamum species and found that the five wild species have higher lignans content (Table 1.6). Interestingly, in the cultivated sesame, sesamin and sesamolin contents are relatively low and almost equal to each other. S. angolense has the highest lignan content of 15.18 mg/g seed. S. latifollum has no sesamolin but has high sesamin of 7.87 mg/g. Meanwhile, the sesamolin content in the other four species is relatively high. Compared with S. indicum, the five wild species has lower sesamin content, even though the total content of sesamin and sesamolin is higher.

Many reports have demonstrated that sesame oil and sesamin have the therapeutic benefits for anti-inflammation in osteoarthritis, antioxidant, antihypertensive, anticancers, lowering blood cholesterol and serum lipid, and neuroprotective effects against hypoxia (Fukuda et al. 1986; Tsuruoka et al. 2005; Srisuthtayanont et al. 2017). Sesamin, sesaminol, and episesamin can specifically inhibit the Δ^5 desaturases which participate in polyunsaturated fatty acid biosynthesis in fungus and rat liver and play an important role in health-promoting effects (Shimizu et al. 1991).

Sesamin is the most important lignans in sesame oil (Kanu et al. 2010). Under normal room temperature and atmospheric pressure, sesamin presents needle-like crystal. The melting point is at 122–123 °C. The boiling point is 504.4 °C at 760 mmHg pressure. Sesame is lipid-soluble. Sesamin can possibly inhibit the catabolism of γ -tocopherol, which results in higher bioavailablility observed in human and animal studies (Ikeda et al. 2002; Sontag and

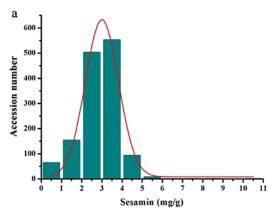
Fig. 1.9 Chemical structure of sesame lignans. *Source* from Ide et al. (2003)

Sesamum species	Material Variety name	Lignans (mg/g seed)		
		Sesamin	Sesamolin	Total
S. indiucm	Yuzhi 11*	2.37	2.23	4.60
S. latifolium	K1*	7.87	0	7.87
S. angolense	K16*	0.54	14.64	15.18
S. calycinum	Ken8*	0.52	5.99	6.51
S. angustifolium	G01*	0.18	12.91	13.09
S. radiatum	G02*	1.42	5.06	6.48

Table 1.6 Lignans content in the seven *Sesamum* species

ND indicates 'Not detected'

*The above sample seeds are harvested in Sanya Experimental station of HSRC, HAAS, China, in 2015 (unpublished data). The sesamin and sesamolin contents are measured by the high-performance liquid chromatography (HPLC) method. (Provided by Haiyang Zhang)



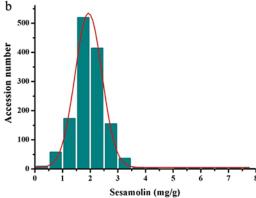


Fig. 1.10 Distribution of sesamin and sesamolin content in the worldwide sesame germplasm. **a** Sesamin content of the 1380 sesame accessions. **b** Sesamolin content of the 1380 sesame accessions (unpublished data). All the

materials are cultivated in Yuanyang experimental station of HSRC, HAAS, China, in 2015. *Source* from Zhang et al. (2019)

Parker 2002). Srisuthtayanont et al. (2017) proved that sesamin increased expression of all the chondroitin sulfate proteoglycans (CSPGs) synthesis genes and suppressed the interleukin-1 beta (IL-1 β) expression in genes and in protein levels.

In addition, sesamin shows antihypertensive and neuroprotective effects against hypoxia or brain damage. As the DOCA-salt hypertensive rats are fed with sesamin, the enhanced vascular O²⁻ production is inhibited. The results indicate that sesamin bears the antioxidative action and contributes to the antihypertensive activity in biomass (Nakano et al. 2002).

Tsuruoka et al. (2005) performed the DNA microarray analysis of the effect of sesamin on the gene expression in rat liver. The results showed that 38 genes were up-regulated of which 16 genes encode proteins possessing a lipid-metabolizing function, and other 16 encode proteins possessing a xenobiotic/endogenous substance metabolizing function. The expression of β -oxidation-associated enzymes in peroxisomes and the auxiliary enzymes required for degradation increased. The expression of acyl-CoA thioesterase genes involved in acyl-CoA hydrolase and very long-chain acyl-CoA thioesterase was also up-regulated. The results suggest