Compendium of Plant Genomes
Series Editor: Chittaranjan Kole

Hongmei Miao · Haiyang Zhang ·
Chittaranjan Kole *Editors*

The Sesame Genome

Springer

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

Springer

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

This book series is dedicated to my wife Phullara and our children Sourav and Devleena

Chittaranjan Kole

Preface to the Series

Genome sequencing has emerged as the leading discipline in the plant sciences coinciding with the start of the new century. For much of the twentieth century, plant geneticists were only successful in delineating putative chromosomal location, function, and changes in genes indirectly through the use of a number of "markers" physically linked to them. These included visible or morphological, cytological, protein, and molecular or DNA markers. Among them, the first DNA marker, the RFLPs, introduced a revolutionary change in plant genetics and breeding in the mid-1980s, mainly because of their infinite number and thus potential to cover maximum chromosomal regions, phenotypic neutrality, absence of epistasis, and codominant nature. An array of other hybridization-based markers, PCR-based markers, and markers based on both facilitated construction of genetic linkage maps, mapping of genes controlling simply inherited traits, and even gene clusters (QTLs) controlling polygenic traits in a large number of model and crop plants. During this period, a number of new mapping populations beyond F_2 were utilized and a number of computer programs were developed for map construction, mapping of genes, and for mapping of polygenic clusters or QTLs. Molecular markers were also used in the studies of evolution and phylogenetic relationship, genetic diversity, DNA fingerprinting, and map-based cloning. Markers tightly linked to the genes were used in crop improvement employing the so-called marker-assisted selection. These strategies of molecular genetic mapping and molecular breeding made a spectacular impact during the last one and a half decades of the twentieth century. But still they remained "indirect" approaches for elucidation and utilization of plant genomes since much of the chromosomes remained unknown and the complete chemical depiction of them was yet to be unraveled.

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

As expected, sequencing of chromosomal regions would have led to too much data to store, characterize, and utilize with the-then available computer software could handle. But the development of information technology made the life of biologists easier by leading to a swift and sweet marriage of biology and informatics, and a new subject was born—bioinformatics.

Thus, the evolution of the concepts, strategies, and tools of sequencing and bioinformatics reinforced the subject of genomics—structural and functional. Today, genome sequencing has 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,

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.

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 — Hongmei Miao
Zhengzhou, China — Haiyang Zhang
New Delhi, India — 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

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

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

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

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

1

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.

1.1 Introduction

Sesame (*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;

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

H. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-98098-0_1

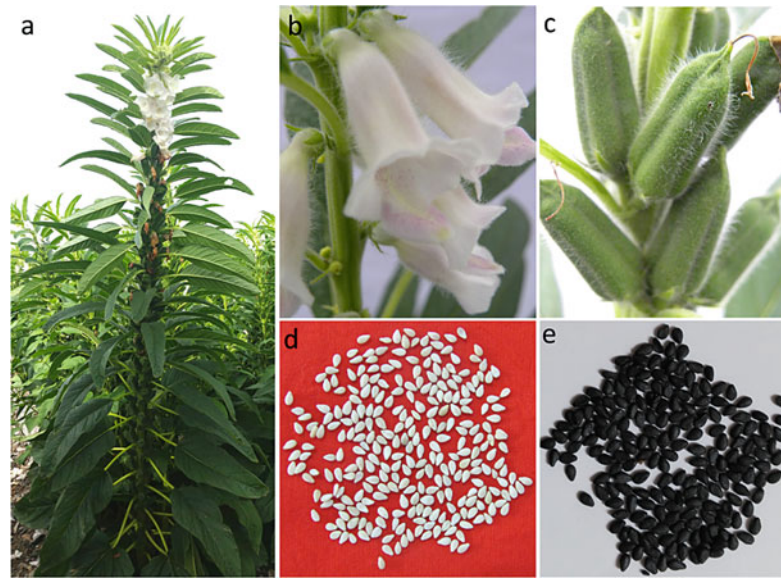


Fig. 1.1 Images of sesame plant, flower, capsule, and seeds. **a** Sesame (Chinese var. Yuzhi Dw607 with short internode length) plant in field. **b** Sesame flowers. **c** Sesame capsules. **d** and **e** White- and black-coated sesame seeds. (Photographs provided by Haiyang Zhang)

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

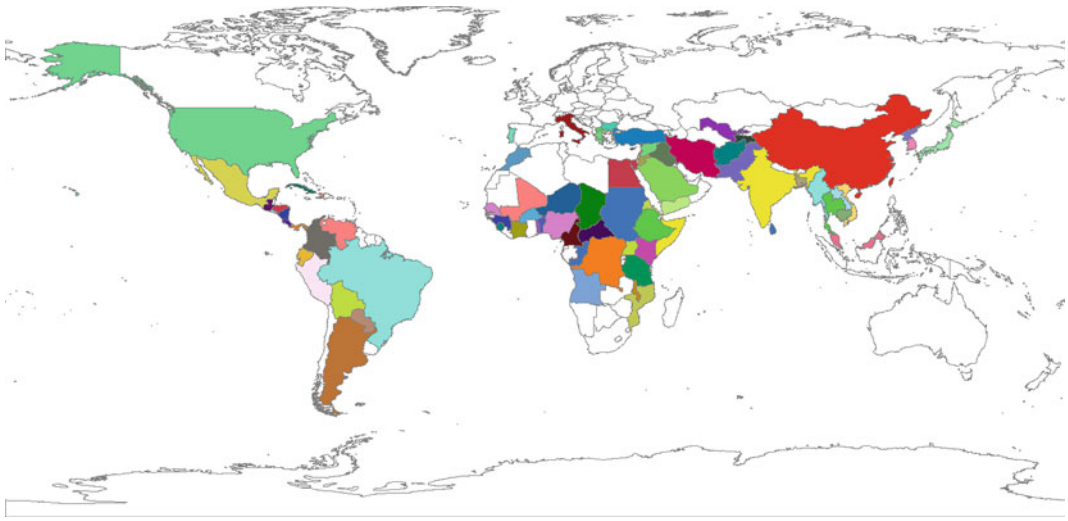


Fig. 1.2 Distribution of the world sesame production. Eighty sesame production countries are shown in color

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

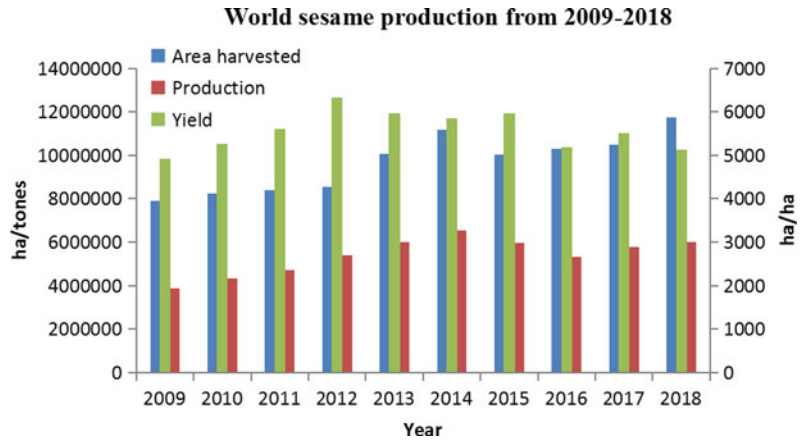


Fig. 1.3 Trends in the world sesame production from 2009 to 2018

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

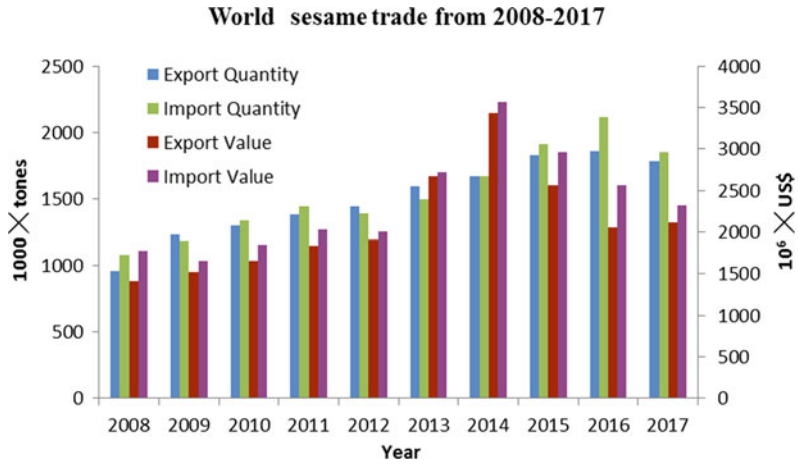


Fig. 1.4 Trends in the world sesame trade from 2008 to 2017

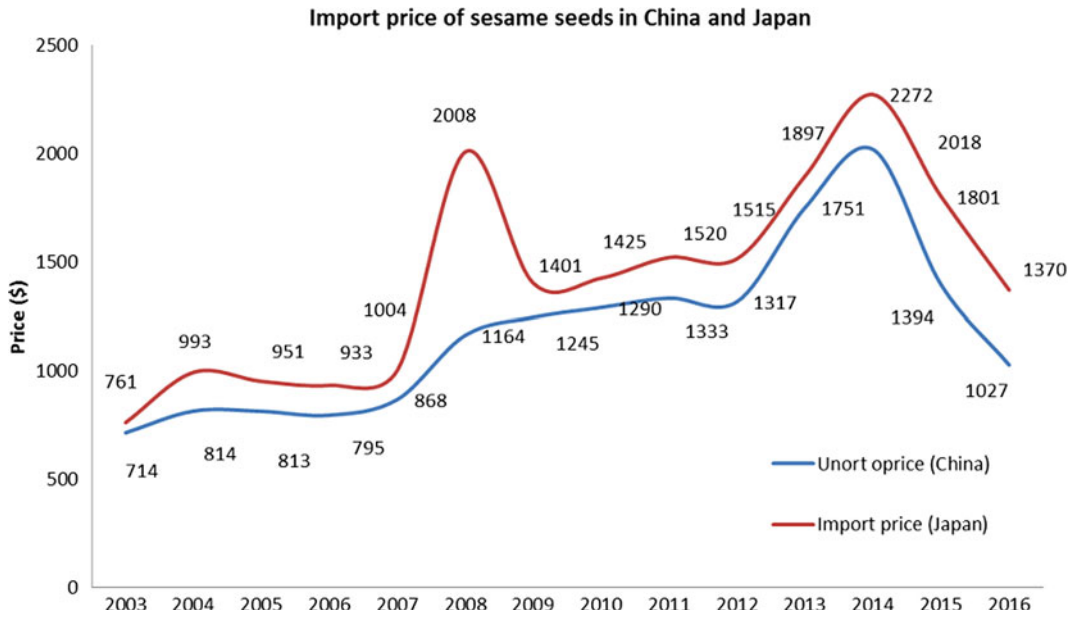


Fig. 1.5 Import price variation of sesame seeds in China and Japan from 2003 to 2016. *Source* from Chinese customs dataset

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

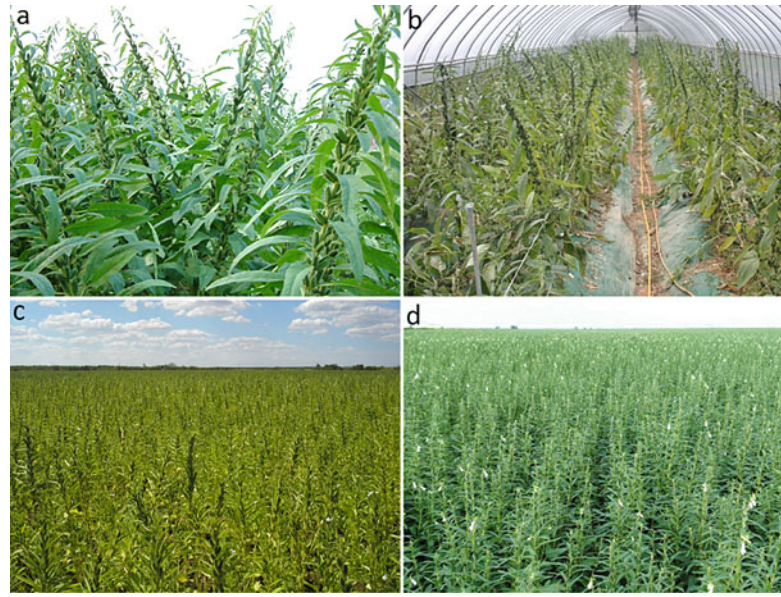


Fig. 1.7 Representative sesame varieties planted in various conditions. **a** Sesame variety Yuzhi11 in China. **b** Sesame variety in South Korea. **c** Sesaco sesame variety in USA. **d** Sesame variety in Sudan (Photographs provided by Haiyang Zhang, Churl Whan Kang, Ray Langham, and Khalafalla Ali, respectively)

Table 1.2 Phases and stages of sesame

Stage/phase	End point of stage	DAP[a]	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	Branches/minor plants stop flowering	53–81	4
Late bloom	90% of plants with no open flowers	82–90	1+
Ripening	Physiological maturity (PM)	91–106	2+
Drying			
Full maturity	All seed mature	107–112	1−
Initial drydown	First dry capsule	113–126	2
Late drydown	Full drydown	137–146	3

Source from Langham (2007)
[a]DAP = days after planting. Data were collected based on S26 in 2004 at Uvalde, Texas, under irrigation condition

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 has 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

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

[a]Vitamin content data are cited from Zhang et al. (2012)
[b]Mineral 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 (B_2) (6.3 mg), niacin (5.0 mg), tocopherol (α, β) (5 mg), thiamine (B_1) (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 B_6. 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) (~40%) and linoleic acid (18:2) (~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
Tyrosine	0.90	0.84
Threonine	0.89	0.86
Methionine	0.84	0.61
Isoleucine	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
Tryptophan	0.21	0.20
Total	20.42	19.33

[a]White 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, and 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^5desaturases 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

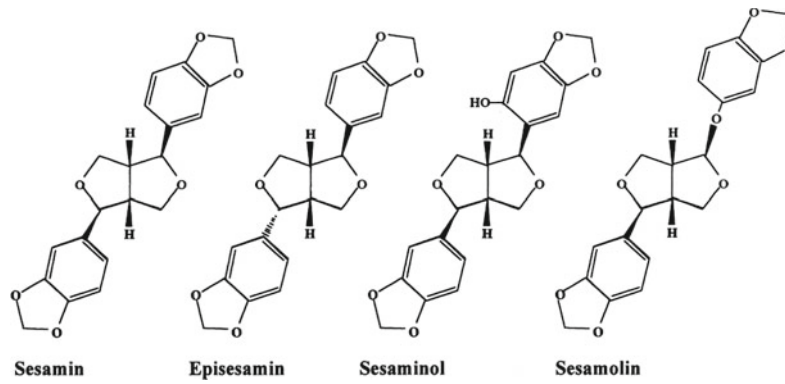


Fig. 1.9 Chemical structure of sesame lignans. *Source* from Ide et al. (2003)

Table 1.6 Lignans content in the seven *Sesamum* species

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

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)

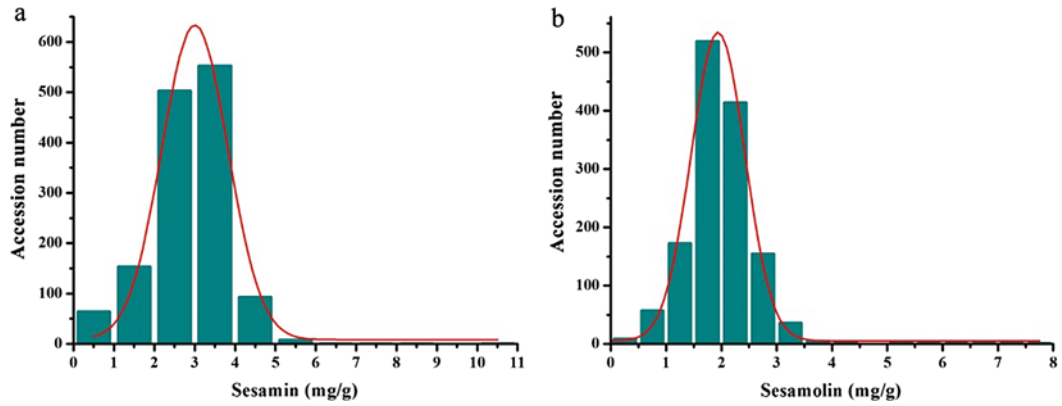


Fig. 1.10 Distribution of sesamin and sesamolin content in the worldwide sesame germplasm. **a** Sesamin content of the 1380 sesame accessions. **b** Sesamolin content of the 1380 sesame accessions (unpublished data). All the materials are cultivated in Yuanyang experimental station of HSRC, HAAS, China, 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-1β) 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 O^{2-} 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

that sesamin regulates the lipids metabolism at the mRNA level. Furthermore, studies indicate that NF-κB, STAT3, JNK, ERK1/2, p38 MAPK, PI3K/AKT, caspase-3, and p53 signaling pathways are critically involved in mediating the anticancer effects of sesamin. Thus sesame can be used for cancer prevention and treatment (Majdalawieh et al. 2017).

Sesamol is an effective antioxidant found mainly in roasted sesame seeds or processed sesame oil. In sesamol molecule, there is a phenolic derivative with a methylenedioxy group. Sesamol can inhibit the excessive production of nitric oxide in the lipopolysaccharide/gamma-interferon stimulated C6 astrocyte cells (Soliman and Mazzio 1998). Studies have shown that sesamol can act as a metabolic regulator and possesses chemopreventive, antimutagenic, and antihepototoxic properties (Kaur and Saini 2000; Kapadia et al. 2002).

Sesaminol is naturally a trace compound in sesame seeds and can be used as a food additive or in medicinal application for the natural antioxidance. Sesaminol could be transformed from sesamolin. The antioxidative activity of sesaminol is relatively equal to those of sesamol and γ-tocopherol. The antioxidative activity of refined unroasted seed oil probably results from sesaminol (Fukuda et al. 1986; Miraj and Kiani 2016).

Sesamolin exists in unprocessed sesame oil and is the source of sesaminol. Sesamolin can be converted to sesamol, sesaminol dimmer, and other lignans, when heated. Sesamolin achieves the anticancer activity by increasing the expression level of NKG2D ligands on Raji cells, which are derived from Burkitt's lymphoma (Table 1.7).

Table 1.7 Physical characteristics and main functions of major sesame lignans

Sesame lignan	Content in seeds (mg/100 g)	Molecular formula	Molecule size	Content in sesame oil (%)	Biological function
Sesamin	77–930	$C_{20}H_{18}O_6$	354	0.40	Lipid and glucose metabolism Hypertension Anti-inflammatory activity by inhibiting delta 5-desaturase Increasing γ-tocopherol levels in plasma and liver of humans Free radical scavenging Protecting liver against ethanol- and carbon tetrachloride-induced damage Inhibiting vascular superoxide production Promoting angiogenesis Providing neuroprotection Bactericide and insecticide activities
Sesamolin	61–530	$C_{20}H_{18}O_7$	370	0.30	Increasing the hepatic mitochondrial and the peroxisomal fatty acid oxidation rate Synergistic for pyrethrum insecticides Inhibiting mutagenesis induced by H_2O_2
Sesaminol	0.3–1.4	$C_{20}H_{18}O_7$	370	0.10	Inhibiting the membrane lipid peroxidation Microsomal peroxidation induced by ADP-Fe^{3+}/NADPH Oxidation of LDL induced by copper ions Increasing the availability of tocopherols Inhibiting oxidative damages in DNA

Source from Moazzami (2006) and Dar and Arumugam (2013)

1.6 Sesame Application

1.6.1 Sesame Nutritional Food Production

Traditionally, sesame seeds are used in Hindu culture as a 'symbol of immortality', and sesame oil is used widely in prayers and burial ceremony. At present, sesame is described as a food and aliment for providing energy, a soothing frame of mind, and retarding aging (Prakash and Naik 2014). The sesame market is on the rise, due to the development of more and more industrialized products for consumption, which has growth of around 15% per year. As the main demand for sesame comes from the food industry, 70% of sesame seeds in most importer countries are used for crushing oil and producing sesame flour and fried seeds. European people use sesame oil as a substitute for olive oil. Japanese use sesame oil for cooking fish.

Sesame seeds are consumed in nature or in products, such as confectionery and bakery. Thus, hulled sesame seeds are also one sort of the important sesame products. The seed coat can be removed via manual, mechanical, physical, and chemical processes. Tahini made from the roasted hulled sesame seeds is a traditional sesame food in Eastern Asia. Meanwhile, sesame seed cake is nutritious and can be used as animal feed and sesame flour. Sesame flour has high protein content. Especially methionine and tryptophan take up 10–12% sesame proteins (Anilakumar et al. 2010).

Moreover, sesame seeds are applied as an ingredient for food production. In Greece, seeds are used in cakes, while in Togo and Africa, the seeds are an ingredient of main soups. Many recipes, such as sesame seed sprouts, hummus, sesame mustard sauce, ginger sesame chicken, sesame pastry, sesame seed sauce, and sesame green beans, also contain sesame seeds as an important ingredient (Anilakumar et al. 2010).

1.6.2 Application as Medicinal Substances

Oxidative stress is associated with many diseases by radical damage, such as atherosclerosis, diabetes mellitus, chronic renal failure, rheumatoid arthritis, and neurodegenerative diseases (Abuja and Albertini 2001). Lipid peroxidation reactions are generally free radical-driven chain reactions, in which one radical can induce the oxidation of a great number of substrate molecules represented by polyunsaturated fatty acids (PUFA). As for the nutritious and antioxidant characters, sesame oil, lignans, lecithin, flavonoids, and other nutritious constitutes extracted from sesame seeds are used as antifungal, antioxidant, cancer preventative drugs, skin softener, and other medicinal and pharmaceutical substances. In the USA medicine market, 'Scivation' company sells sesame seed ligans soft gels applied for fat loss, reducing cholesterol to support liver and kidney health, and maintaining healthy lipid profile. Of 540 mg sesame seed extract, there are 81 mg (15%) sesamin and 27 mg (5%) sesamolin.

1.6.3 Industry Processing

In addition, sesame can be used for industry for specific physical characters. For example, sesame flowers can be used to produce perfumes and cologne. Myristic acid has been extracted from sesame oil and used for cosmetics and soap manufacturing. Sesamolin has insecticidal properties and has been used as a synergist for pyrethrum insecticides (Morris 2002). Moreover, sesame oil has been used as a solvent, oleaginous vehicle for drugs, skin softener, and the manufacture of margarine and soap. Chlorosesamone has been obtained from sesame roots and proved with antifungal activity (Hasan et al. 2000).

Furthermore, sesame cake and rejected materials can be used as fertilizes. The stems of sesame plants can produce organic fertilizers to improve the soil nutrition. It is also worthy to mention that new further processing products of sesame are creating new consumption market for sesame industry in recent years. As to new sesame products related with proteins and lignans, new processing techniques always attract higher additional value and more producers' attention. We therefore believe that sesame industry is vigorous with the sustainable development tendency.

1.7 Academic Importance

Sesame takes both the important economic position and academic position during the development of the world oilseed crops industry. Compared with oilseed rape (*Brassica napus*), soybean (*Glycine max*), peanut (*Arachis hypogaea*), sunflower (*Helianthus annuus*), and other oilseed crops, sesame seeds contain high fatty acids and antioxidants content. Clarifying the regulatory mechanisms underlying oil biosynthesis, fatty acid and storage protein composition, and the secondary metabolic processes of antioxidants in sesame will certainly explore the energy substance synthesis patterns in higher plants and oilseed crops (Suh et al. 2003; Ashri 2006; Wei et al. 2012; Zhang et al. 2013).

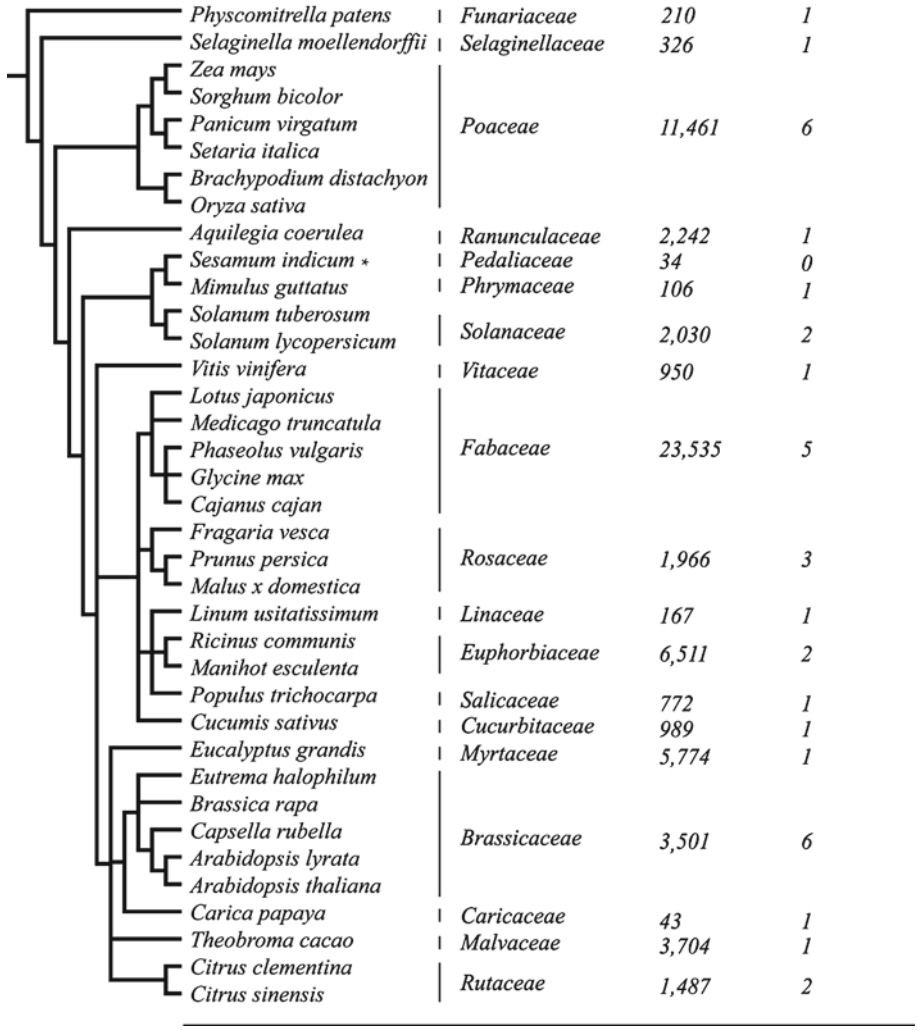
Sesame is a very ideal material for genetic studies. Sesame grows fast with the high productive character. The life cycle stage is about 71–99 days, while the flowering stage varies from 28 to 65 days (Kang et al. 2006). Of the 806 germplasm accessions cultured in Pingyu experimental station (Pingyu, China) in 2017, the capsule number per plant varies from 6 to 165, and the average seed number per capsule varies from 20 to 173 (unpublished data, provided by Zhang Haiyang). Thus, obtaining the offspring of hybridization population is very easy for sesame.

In addition, sesame has high adaptability for tropical climates. The tolerance of drought and high temperatures make sesame survival successful in low input and arid lands. Sesame is susceptible to fungi diseases (such as *Fusarium* wilt and stem charcoal rot diseases) and waterlogging, while many wild species present high resistance to biotic and abiotic stresses (Joshi 1961; Nimmakayala et al. 2011; Zhang et al. 2012). Comparison of the genetic and genomic differences between sesame and the wild *Sesamum* species will determine the stress response mechanism of disease resistance in sesame. Thus, the *Sesamum* genus can be used as the model plant to investigate the resistance and regulation processes for biotic and abiotic stresses in nature.

More importantly, sesame is an ancient crop, and the evolutionary position is important. Chloroplast genomics data indicate that *S. indicum* in *Sesamum* (Pedaliaceae) is closely related to the family Solanaceae (Fig. 1.11). The sesame genome is small to only 354 Mb according to the *k*-mer analysis results (Zhang et al. 2013). The chromosome structure is relatively simple and presents the diploid characters (Zhao et al. 2017). Thus, clarifying the genome evolution of *S. indicum* and other wild species in *Sesamum* will supply precious information for revealing the possible evolution pattern of the ancestors of eudicot plants. Therefore, sesame is regarded as a model oilseed crop for genetic studies of key agronomic traits and genomic evolution analysis within endicot plants (Zhang et al. 2019).

1.8 Conclusion

Sesame is an important oilseed crop with the world production of 6.11 million tons annually. Sesame as a model of oilseed species offers a great opportunity for genetics analysis of vegetable oil biosynthesis and metabolism. With the fulfilling of the Sesame Genome Project, sesame can supply an example for exploring the



Genome-sequenced Species	Family	Species number	Available genome number
Physcomitrella patens	*Funariaceae*	*210*	*1*
Selaginella moellendorffii	*Selaginellaceae*	*326*	*1*
Zea mays	*Poaceae*	*11,461*	*6*
Sorghum bicolor			
Panicum virgatum			
Setaria italica			
Brachypodium distachyon			
Oryza sativa			
Aquilegia coerulea	*Ranunculaceae*	*2,242*	*1*
Sesamum indicum *	*Pedaliaceae*	*34*	*0*
Mimulus guttatus	*Phrymaceae*	*106*	*1*
Solanum tuberosum	*Solanaceae*	*2,030*	*2*
Solanum lycopersicum			
Vitis vinifera	*Vitaceae*	*950*	*1*
Lotus japonicus	*Fabaceae*	*23,535*	*5*
Medicago truncatula			
Phaseolus vulgaris			
Glycine max			
Cajanus cajan			
Fragaria vesca	*Rosaceae*	*1,966*	*3*
Prunus persica			
Malus x domestica			
Linum usitatissimum	*Linaceae*	*167*	*1*
Ricinus communis	*Euphorbiaceae*	*6,511*	*2*
Manihot esculenta			
Populus trichocarpa	*Salicaceae*	*772*	*1*
Cucumis sativus	*Cucurbitaceae*	*989*	*1*
Eucalyptus grandis	*Myrtaceae*	*5,774*	*1*
Eutrema halophilum	*Brassicaceae*	*3,501*	*6*
Brassica rapa			
Capsella rubella			
Arabidopsis lyrata			
Arabidopsis thaliana			
Carica papaya	*Caricaceae*	*43*	*1*
Theobroma cacao	*Malvaceae*	*3,704*	*1*
Citrus clementina	*Rutaceae*	*1,487*	*2*
Citrus sinensis			

Fig. 1.11 Phylogenetic positions of sesame and the 36 land plants with available genome sequences *Source* from Zhang et al. (2013).

molecular mechanisms related with growth and development and resistance responses to biotic and abiotic stresses for other crops.

References

Abou-Gharbia HA, Shahidi F, Adel A, Shehata M, Youssef MM (1997) Effect of processing on oxidative stability of sesame oil extracted from intact and dehulled seeds. J Am Oil Chem Soc 74:215–221

Abou-Gharbia HA, Adel A, Shahidi F (2000) Effect of processing on oxidative stability and lipid classes of sesame oil. Food Res Int 33:331–340

Abuja PM, Albertini R (2001) Methods for monitoring oxidative stress, lipid peroxidation and oxidation resistance of lipoproteins. Clin Chim Acta 306(1–2):1–17

Anilakumar KR, Pal A, Khanum F, Bawa AS (2010) Nutritional, medicinal and industrial uses of sesame (*Sesamum indicum* L.) seeds—an overview. Agriculturae Conspectus Scientificus 75(4):159–168

Ashri A (1998) Sesame breeding. In: Janick J (ed) Plant breeding reviews. Wiley, Israel, pp 179–228

Ashri A (2006) Sesame (*Sesamum indicum* L.). In: Singh RJ (ed) Genetic resources, chromosome engineering, and crop improvement. CRC Press, Boca Raton, FL, pp 231–280

Baydar H (2005) Breeding for the improvement of the ideal plant type of sesame. Plant Breed 124(3):263–267

Bedigian D, Harlan JR (1986) Evidence for cultivation of sesame in the ancient world. Econ Bot 40:137–154

Bedigian D, Seigler DS, Harlan JR (1985) Sesamin, sesamolin and the origin of sesame. Biochem Syst Ecol 13(2):133–139

Dar AA, Arumugam N (2013) Lignans of sesame: purification methods, biological activities and biosynthesis—a review. Bioorg Chem 50:1–10

Fukuda Y, Nagata M, Osawa T, Namkini M (1986) Contribution of lignan analogues to antioxidative activity of refined unroasted sesame seed oil. J Am Oil Chem Soc 63(8):1027–1031

Gao TM, Wei SL, Zhang HY, Chun-Ming LI, Zheng YZ et al (2011) Analysis of the nutritional components of white sesame kernel. Acta Nutrimenta Sin 53 (12):565–581

Halliwell B (1990) How to characterize a biological antioxidant. Free Radical Res Commun 9(1):1–32

Hasan A, Begum S, Furumoto T, Fukui H (2000) A new chlorinated red naphthoquinone from roots of *Sesamum indicum*. Biosci Biotechnol Biochem 64:873–874

Ide T, Masayo K, Yoko T, Kazuki S, Nobuhiro F, et al (2003) . Sesamin, a Sesame Lignan, as a Potent Serum Lipid-Lowering Food Component. Japan Agri Res Quarterly 37(3):151–158

Ikeda S, Tohyama T, Yamashita K (2002) Dietary sesame seed and its lignans inhibit 2,7,8-trimethyl-2(2'-carboxyethyl)-6-hydroxychroman excretion into urine of rats fed γ-Tocopherol. J Nutr 132(5):961–966

Jimoh WA, Aroyehun HT (2011) Evaluation of cooked and mechanically defatted sesame (*Sesamum indicum*) seed meal as a replacer for soybean meal in the diet of African catfish (*Clarias gariepinus*). Turk J Fish Aquat Sci 11(2):185–190

Joshi AB (1961) Sesamum. Indian central oilseed committee Hyderabad, India, p 109

Kamaleldin A, Appelqvist LA (1996) Aldehydic acids in frying oils: formation, toxicological significance and analysis. Grasas Aceites 47(5):342–348

Kang CW, Seo K, Sung L, Prem M, Toby H et al (2006) Selection of a core collection of Korean sesame germplasm by a stepwise clustering method. Breed Sci 56:85–91

Kanu PJ, Bahsoon JZ, Kanu JB, Kandeh JBA (2010) Nutraceutical importance of sesame seed and oil: a review of the contribution of their lignans. Sierra Leone J Biomed Res 2(1):4–16

Kapadia GJ, Azuine MA, Tokuda H, Takasaki M, Mukainaka T et al (2002) Chemopreventive effect of resveratrol, sesamol, sesame oil and sunflower oil in the Epstein-Barr virus early antigen activation assay and the mouse skin two-stage carcinogenesis. Pharmacol Res 45(6):499–505

Kaur IP, Saini A (2000) Sesamol exhibits antimutagenic activity against oxygen species mediated mutagenicity. Mutat Res-Genet Toxicol Environ Mutagen 470 (1):71–76

Kiani S, Zakerhamidi MS, Tajalli H (2016) Hydrogen bonding intermolecular effect on electro-optical response of doped 6PCH nematic liquid crystal with some azo dyes. Opt Mater 55:121–129

Kobayashi T (1986) Some genetic patterns for the main characteristics of the cultivated sesame. In: Fernandez MJ (ed) Sesame and safflower newsletter. CIDA, Cordova, Spain, pp 23–32

Langham DR (2007) Phenology of sesame. In: Janick J, Whipkey A (eds) Issues in new crops and new uses. ASHS Press, Alexandria, VA, pp 144–182

Langham DR (2008) Growth and development of sesame. American sesame grower association, San Antonio, TX, p 44

Li C, Miao H, Wei L, Zhang T, Han X, Zhang H (2014) Association mapping of seed oil and protein content in *Sesamum indicum* L. using SSR markers. PLoS One 9 (8):e105757

Majdalawieh AF, Fayyad MW, Nasrallah GK (2017) Anti-cancer properties and mechanisms of action of thymoquinone, the major active ingredient of *Nigella sativa*. Crit Rev Food Sci Nutr 57(18):3911–3928

Makinde FM, Akinoso R (2013) Nutrient composition and effect of processing treatments on anti nutritional factors of Nigerian sesame (*Sesamum indicum* Linn) cultivars. Int Food Res J 20(5):2293–2300

Miraj S, Kiani S (2016) Bioactivity of *Sesamum indicum*: a review study. Der Pharmacia Lett 8(6):328–334

Moazzami AA (2006) Sesame seed lignans: diversity, human metabolism and bioactivities. A doctoral thesis submitted to the Swedish University of agricultural sciences, Uppsala, Sweden

Morris JB (2002) Food, industrial nutraceutical uses of sesame genetic resources. In: Janick, Whipkey A (eds) Trends in new crops and new uses. ASDHS Press, Alexandria, USA, pp 153–156

Nakano D, Itoh C, Takaoka M, Kiso Y, Tanaka T et al (2002) Antihypertensive effect of sesamin. IV. Inhibition of vascular superoxide production by sesamin. Biol Pharm Bull 25(9):1247–1249

Nimmakayala P, Perumal R, Mulpuri S, Reddy UK (2011) Sesamum. In: Kole C (ed) Wild crop relatives: genomic and breeding resources, vol Oilseeds. Springer, Berlin Heidelberg, pp 261–273

Prakash K, Naik SN (2014) Bioactive constituents as a potential agent in sesame for functional and nutritional application. J Bioresour Eng Technol 1:48–66

Ram R, Catlin D, Romero J, Cowley C (1990) Sesame: new approaches for crop improvement. In: Janick J, Simon JE (eds) Advances in new crops. Timber Press, Portland, pp 225–228

Sene B, Sarr F, Sow MS, Diouf D, Niang M (2017) Physico-chemical composition of the sesame variety

(*Sesamum indicum* L.) and characterization of its derived products (seeds, oil and oilcake) in Senegal. Food Sci Qual Manag 65:5–10

Shimizu S, Akimoto K, Shinmen Y, Kawashima H, Sugano M et al (1991) Sesamin is a potent and specific inhibitor of Δ5 desaturase in polyunsaturated fatty acid biosynthesis. Lipids 26(7):512–516

Shittu LAJ, Bankole M, Ahmed T, Bankole MN, Shittu RK et al (2007) Antibacterial and antifungal activities of essential oils of crude extracts of *Sesame radiatum* against some common pathogenic microorganisms. Iran J Pharmacol Ther 6(6):165–170

Smith DE, Salerno JW (1992) Selective growth inhibition of a human malignant melanoma cell line by sesame oil *in vitro*. Prostaglandins Leukot Essent Fatty Acids 46(2):145–150

Soliman KF, Mazzio EA (1998) *In vitro* attenuation of nitric oxide production in C6 astrocyte cell culture by various dietary compounds. Proc Soc Exp Biol Med 218(4):390–397

Sontag TJ, Parker RS (2002) Cytochrome P450 omega-hydroxylase pathway of tocopherol catabolism. Novel mechanism of regulation of vitamin E status. J Biol Chem 277(28):25290–25296

Srisuthtayanont W, Pruksakorn D, Kongtawelert P, Pothacharoen P (2017) Effects of sesamin on chondroitin sulfate proteoglycan synthesis induced by interleukin-1beta in human chondrocytes. BMC Complement Altern Med 17(1):286

Suh MC, Kim MJ, Hur CG, Bae JM, Park YI et al (2003) Comparative analysis of expressed sequence tags from *Sesamum indicum* and *Arabidopsis thaliana* developing seeds. Plant Mol Biol 52:1107–1123

Suwimol S, Wiroj JR, Walin WK, Natchapon J, Pathamaporn H et al (2012) Effects of sesame seeds consumption on serum cholesterol and oxidative status in hypercholesterolemia. Food Public Health 2(6):193–196

Tsuruoka N, Kidokoro A, Matsumoto I, Abe K, Kiso Y (2005) Modulating effect of sesamin, a functional lignan in sesame seeds, on the transcription levels of lipid- and alcohol-metabolizing enzymes in rat liver: a DNA microarray study. Biosci Biotechnol Biochem 69(1):179–188

Uzun B, Arslan C, Furat S (2008) Variation in fatty acid composition, oil content and oil yield in a germplasm collection of sesame (*Sesamum indicum* L.). J Am Oil Chem Soc 85:1135–1142

Wei L, Miao H, Zhang H (2012) Transcriptomic analysis of sesame development. Sci Agric Sin 45(7):1246–1256

Wei L, Miao H, Li C, Zhang H (2016) Establishment of NMR method for rapid and nondestructive detection of oil content in sesame seeds. Jiang Agric Sci 44 (8):386–388

Zhang H, Wang X, Wang H, Wei S (2012) Sesame production technology. Henan people press, Zhengzhou, China, p 203

Zhang H, Miao H, Wang L, Qu L, Liu H et al (2013) Genome sequencing of the important oilseed crop *Sesamum indicum* L. Genome Biol 14(1):401

Zhang H, Miao H, Ju M (2019) Potential for adaptation to climate change through genomic breeding in sesame. In: Kole C (ed) Genomic designing of climate-smart oilseed crops. Springer, Cham, pp 374–376

Zhao R, Miao H, Song W, Chen C, Zhang H (2017) Identification of sesame (*Sesamum indicum* L.) chromosomes using the BAC-FISH system. Plant Biol 20 (1):85–92

Botanical Descriptions of Sesame

2

Hongmei Miao, D. Ray Langham, and Haiyang Zhang

Abstract

Sesame is distributed widely in the tropical and subtropical regions of the world. The cultivation history of sesame dates back to over 5000 years ago. In this section, the botanical features and growth characteristics of sesame including root, stem, leaf, flower, capsule, and seed are discussed. The overview summarizes the species-specific morphological characteristics of sesame and provides the fundamentals for sesame genetics and genomics research. In addition, the wild species are also described.

2.1 Origin of *Sesamum indicum*

Sesame was named '*Sesamum indicum*' by Carl Linnaeus for the first time in 1753 (Bedigian 2010a). The name '*Sesamum orientale*' precedes *S. indicum*, and many scientists have argued that *S. orientale* should be used, and some papers have used it. However, the debate was settled by the Committee for Spermatophyta at the 2005 International Botanical Congress that approved the name *Sesamum indicum* L. in the International Code of Botanical Nomenclature (ICBN) (Brummitt 2005).

S. indicum is the sole widely cultivated species in the *Sesamum* genus (Joshi 1961; Weiss 1971; Zhang et al. 2012), although there are minor uses of a wild species *S. radiatum*. Sesame is widely grown in Asia, Africa, and America. To ascertain the original place and the expansion history is difficult. According to Index Kewens and Supplements, about 29 of the 36 *Sesamum* species have been found in Africa (Kobayashi 1991). The existence of an overwhelming number of wild *Sesamum* species in Africa suggests that this continent should be the place of origin of sesame (Joshi 1961; Weiss 1983; Sharma et al. 2014). However, Asia is rich in diversity of cultivated sesame. Fuller (2003) inferred that sesame originated from the northwestern South Asia by the time of the Harappan civilization, based on the archaeological findings of charred sesame seeds. There is a species (*S. malabaricum* Burm.) in India, which can be crossed with *S. indicum* producing fertile seeds (Bedigian 2010a). The analysis of the two chloroplast DNA regions (*ndh*F and *trn*LF) determined that *S. indicum* is closely related with *S. malabaricum* (Bedigian 2010b). Nanthakumar et al. (2000) studied the genetic relationships among seven cultivated and wild *Sesamum* species using isozymes and random amplified polymorphic DNA (RAPD) markers. The data showed that *S.*

H. Miao · H. Zhang (✉)
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. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-98098-0_2

malabaricum and *S. radiatum* were most closely related to cultivated *S. indicum*. Hiremath and Patil (1999) regarded *S. mulayanum* as the wild progenitor of *S. indicum*, according to the morphological, genetic, and cytogenetic analyses. Thus sesame is believed to be domesticated in the Indian subcontinent (the western Indian peninsula and parts of Pakistan) (Bedigian 1984, 1985, 1988, 2000; Bedigian et al. 1985; Bedigian and Harlan1986; Powell 1991; Zohary and Hopf 1994; Bhat et al. 1999; Hiremath and Patil 1999; Fuller 2000; Fuller and Madella 2000). However, further genomic works should be done to confirm the evolution history and the exact progenitor of *S. indicum*.

2.2 Botanical Traits of the Cultivated Sesame

S. indicum is an erect annual herb. According to the descriptions of botanists, *S. indicum* presents the specific botanical traits as follows: High, simple, or branched. Stem obtusely quandrangular, sulcate, finely pubescent to glabrous, rarely pilose, usually more or less glandular. Leaves very variable, often heteromorphic, opposite or alternate; lower leaves long, petiolate, ovate or ovate-lanceolate, 3-lobed, 3-partite or 3-foliolate, rounded or obtuse at the base, acute at the apex, margins often dentate; upper leaves more shortly petiolate, narrower, oblong-lanceolate to linear-lanceolate, usually entire, and narrowly cuneate at the base; all leaves thinly pubescent and more or less glandular, becoming glabrous; leaves light green, paler and more gray below. Flowers white, pink, or mauve-pink with darker markings. Calyx persistent; lobes oblong, pubescent. Ovary slightly compressed, more or less rounded at the apex, pilose; filaments and style white, anthers light brown; pollen white. Nectaries are yellow. Capsule oblong quadrangular, slightly compressed, deeply 4-grooved, rounded at the base and apex, then rather abruptly and shortly beaked at the apex. Seeds more or less horizontal in the capsule, not winged, black, brown, or white, with faces smooth or rarely slightly veined, never rugulose or reticulate (Bedigian 2010a).

2.2.1 Root

Sesame has tap root system comprising the principal tap root component and fibrous roots with numerous root hairs (Fig. 2.1a–c). Since the roots can penetrate into deeper soil layers to find moisture and nutrients, sesame is tolerant to drought and infertility, and has the high yield potential without irrigation. The wild species *S. radiatum* has stronger tap root with more stout fibrous roots (Fig. 2.1d). Correspondingly, *S. radiatum* has higher tolerance to biotic and abiotic stresses, compared with the cultivated sesame and some other wild species. When sesame is attacked by *Fusarium* wilt disease, the root and the above-ground part do not grow well and often the plant will die (Fig. 2.1e, f).

Sesame roots grow rapidly in the vegetative phase. In clay soil of Sudan, sesame roots penetrate into the soil layer of 25 cm in 10 days, and 115 cm in 50 days (Weiss 2000). Generally, in the proper soil conditions, the roots may be as deep as the plants are tall (Langham 2008). The growth status of root reflects that of the above-ground part in sesame.

Langham (2018a) states that the root distribution is affected by the moisture. If there are rains or irrigations soon after planting, more fibrous roots form in the upper 30 cm of soil with shorter tap toots. The shallow root is susceptible to drought and heavy rains. In the US, the optimum situation for sesame is to plant sesame seeds into soil moisture and then not to supply additional moisture for about 30 days to force the root down further into the soil.

2.2.2 Stem

Sesame is a crop with a single stem or branch. Bedigian (2010a) described that the stem of cultivated sesame is firm, square with ribs at each corner, and furrowed at the midpoint of each side (Fig. 2.2). The stem color is classified into pale green (or yellow), green, green with purple stripes, and purple (Fig. 2.2a–d). The pubescence length and density on the stem vary in different genotypes. Some stems are glabrous, while some

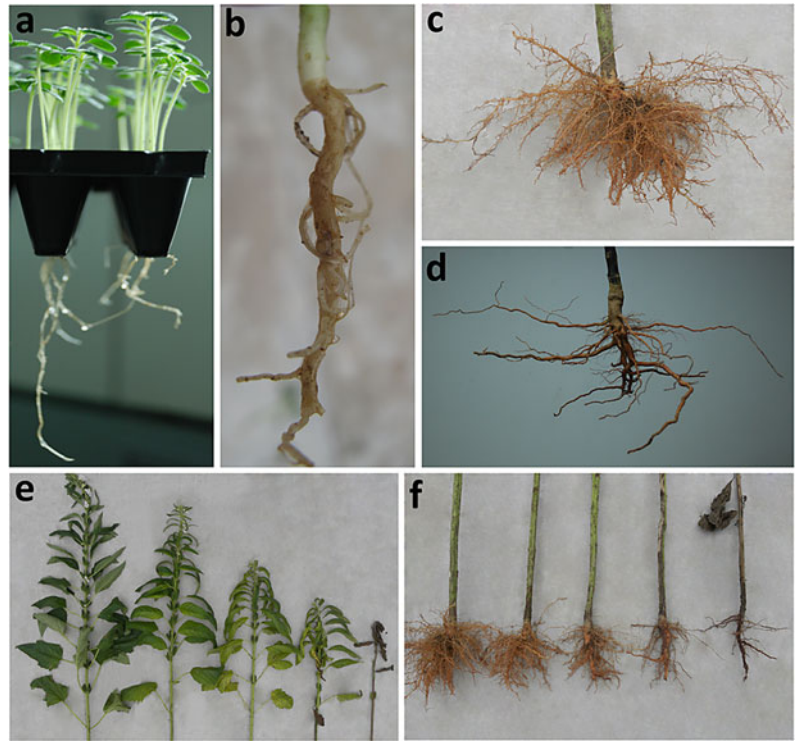


Fig. 2.1 Morphology of sesame root. **a** Root of seedling cultured in nutritional pots. **b** Root of seedling with 3 pairs of leaves in field. **c** Root of adult plant in field. **d** Root of wild species *S. radiatum* with high tolerance to biotic and abiotic stresses. **e** Sesame plants with 0–4 ratings of *Fusarium* wilt symptom, respectively from left to right. **f** Sesame root with 0–4 ratings of *Fusarium* wilt symptom, respectively from left to right. (Photographs provided by Hongmei Miao)

have considerable pubescence (Fig. 2.2e, f). For some sesame mutants and several wild species, the stem is curly or prostrate, and the plant is not erect (Fig. 2.2g, h). In addition, some tetracarpellate genotypes have an association with fasciation (Fig. 2.2i). Only a few flowers and capsules form in the oblate stem (Langham 2007; Bedigian 2010a).

Sesame is a tall plant with relatively short life cycle (about 80–90 days in China). The plant grows slowly in the first 30 days, but then accelerates in the late vegetative and reproductive phases. Histological section observations indicate that the tissue structure of sesame stem includes epidermis, cortex, and vascular cylinder individually from outer to inner layers (Fig. 2.3a, b show a horizontal view, and Fig. 2.3c, d shows a vertical view). In vascular cylinder, there are vascular cells, phloem, xylem, and pith from the outer to the inner tissues. Continuous division of the cambium cells forms new cell layers of phloem and xylem, and result in thickening and lengthening of the stem. When a sesame plant is infected by fungi (such as *Fusarium oxysporum* f. sp. *sesami*, or *Macrophomina phaseolina* (Tassi) Goid.), the stem tissues get brown and gradually lose the transport capability of water and nutrition (partially shown in Fig. 2.4).

There are four key stem-related measurements: plant height, height to the first capsule node, capsule zone length, and tip length without a capsule. The statistics data of stem-related traits in the 763 worldwide sesame accessions were collected in 2017 (Fig. 2.5) (unpublished data, Haiyang Zhang).

2.2.2.1 Plant Height

The height of sesame plant usually is measured from the first leaf node to the highest capsule with viable seeds. Some breeders measure it from the ground to the highest capsule. The plant height varies significantly in the worldwide germplasm. Of the 763 sesame germplasm accessions planted in Pingyu experimental station, China in 2017, plant height varied from 71.67 to 200.00 cm with a normal distribution (Fig. 2.5a) (unpublished data, Haiyang Zhang). More than 50% of the uniculm varieties were

Fig. 2.2 Morphology of sesame stem and type. Stem color, morphology, pubescence density, and stem strength vary in sesame germplasm. **a** Yellowish stem. **b** Green stem. **c** Green stem with purple stripes. **d** Purple stem. **e** Glabrous stem with no pubescence. **f** Pubescent stem. **g** Twisted stem of a mutant. **h** Procumbent stem of a wild *Sesamum* species. **i** Stem fasciation. (Photographs supplied by Haiyang Zhang)

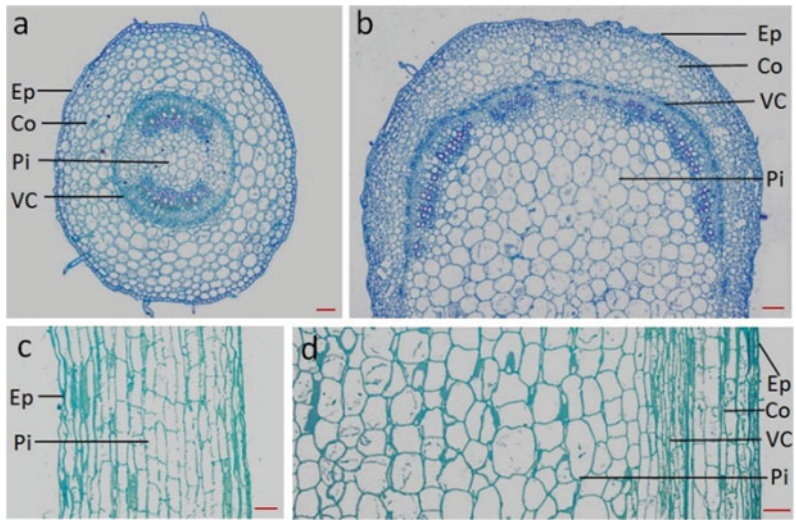


Fig. 2.3 Histological section observation of sesame stem during seedling stage. **a** Horizontal section of sesame stem. **b** The structure of the horizontal stem. **c** Vertical section of sesame stem. **d** The structure of the vertical stem. Chinese variety Yuzhi 11 was used in the study. Ep: epidermis. Co: cortex. VC: vascular cylinder. Pi: pith. Bar = 100 μm. (Photographs supplied by Hongmei Miao)

Fig. 2.4 Sesame stem section of healthy and diseased plants infected by *Fusarium oxysporum* f. sp. *sesami* (FOS). **a** Vertical cut view of left: healthy stem and right: FOS infected stem. **b** Healthy stem cross-section of seedling. **c** FOS infected stem cross-section of seedling. **d** FOS infected stem of adult plant. **e** FOS infected stem of drought plant

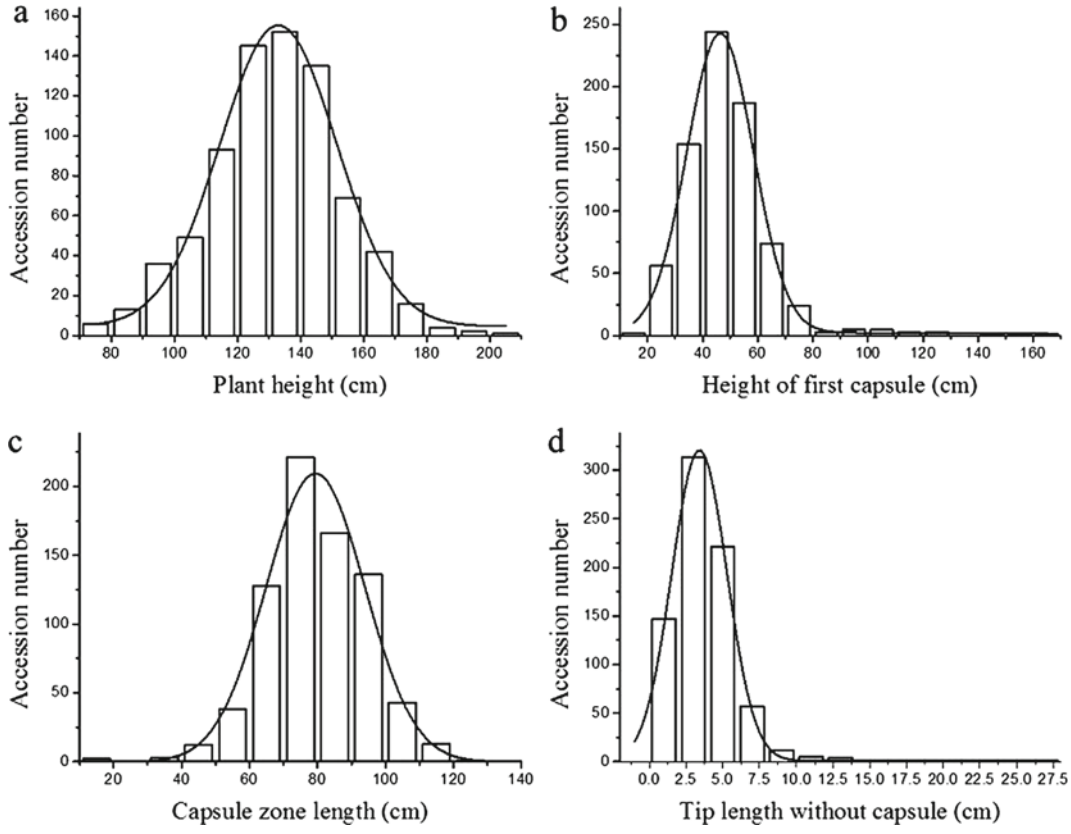


Fig. 2.5 Distribution of values of stem-related traits of 763 worldwide sesame germplasm accessions. **a** Plant height. **b** Height to the first capsule. **c** Capsule zone length. **d** Tip length without a capsule. (Figures supplied by Hongmei Miao)

120–150 cm high. Langham (2019) reported the international current plant height range is 56–311 cm. (There are shorter lines in a drought, but the range includes only lines that were grown in a normal year).

To our knowledge, the highest world record of the plant height of the cultivated sesameis created by a Chinese farmerin 2015 (Zhumadian net, http://news.dahe.cn/2015/09-16/105650582.html). In the usual field size of 0.01 hectares, the average height of the Chinese sesame variety is above 3 m. The highest capsule node number ina uniculm plantis 59 (Fig. 2.6a). Chinese sesame scientists have developed a new sesame variety Yuzhi Dw607 with short internode length. Cultivated in Xinjiang extension station in 2016, the plant height was 1.7 m and had 62 capsule nodeson the uniculm stem (Fig. 2.6b).

The growth speed of the sesame plant varies with the development stages. As shown in Fig. 2.7, the plant height of the cultivated sesame (varieties from Japan and USA, respectively) increases in an 'S' style in the whole cycle. In the seedling stage, the plant height increases slowly. During budding and flowering stages from about 38 days after planting (DAP) to 70 DAP, the height increases significantly (Langham 2007; Sintim and Yeboah-Badu 2010).

Plant height is a complicated trait in sesame. Culp (1960b) found that plant height is controlled by 3–10 factors, with tall plants being dominant. In most cases, transferring the desirable genes controlling the short height to tall genotypes or vice versa usually is time-consuming.

There are some dwarf accessions with low height or short internode length in sesame germplasm from China, Republic of Korea, and Japan. Langham (2011a) studied the line Suwon 128 obtained from Kang CW of the Republic of Korea and found the Suwon 128 is pygmy controlled by a recessive allele. The allele was incorporated into a variety Sesaco 70 with a height of 85 cm with 29 capsule node pairs. The fewer capsule nodes are compensated by the ability to plant in higher plant populations. In 2009, a dwarf line *dw607* was developed from var. Yuzhi 11 using ethylmethane sulfonate (EMS) mutagenesis method by HSRC, HAAS, China (Fig. 2.8). The internode length decreased from 6.0 to 8.4 cm to about 3.5–4.0 cm resulting in a plant height reduction from 170 cm to about 110 cm (data not shown). The height of the first capsule is also lower. Compared with Yuzhi 11, *dw607* has a similar high yield potential but higher resistance to logging. Genetic analysis

Fig. 2.6 Plant height of sesame **a** Chinese domestic sesame variety cultivated in Xiping county, Henan, China by a sesame farmer in 2015 (photo cited from http://news.dahe.cn/2015/09-16/105650582.html). **b** Yuzhi Dw607 with short internode length cultivated in Jinghe county, Xinjiang, China in 2016 (Photographs supplied by Hongmei Miao)

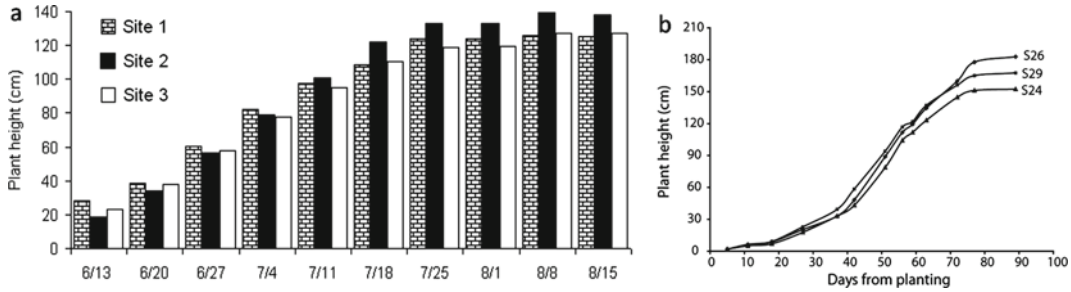


Fig. 2.7 Plant height variation of the cultivated sesame during growth and development. **a** Performance of 3 cultivars adapted from Japan planted in 3 sites. (Modified from Sintim and Yeboah-Badu 2010). **b** The height over time of 3 USA varieties. (Modified from Langham 2007)

Fig. 2.8 Phenotype comparison of dwarf mutant *dw607* with short internode length and the wild type Yuzhi 11. **a** Plant height comparison of Yuzhi 11 (left) and *dw607* (right) at the final flowering stage. **b** Seedlings of Yuzhi 11 (left) and *dw607* (right) at 3 days after germination; **c** Internode length comparison of Yuzhi 11 (left) and *dw607* (right). (Photographs supplied by Haiyang Zhang)

proves that the internode length in *dw607* and Yuzhi 11 is controlled by single gene alleles. To date, dwarf variety Yuzhi dw607 is in production with very good farm yields.

Plant height of sesame can be influenced by the amount of moisture, heat, fertility, light, and population. In continuous cultivation, the plant height and the height to the first capsule node of Sesaco 26 varied under different environments (Langham 2007). In general, the plant grows taller under high moisture and fertility conditions. In a high population with inadequate sunlight, the height of the first capsule node always is high and can reach to more than 1 m.

Plant height is positively correlated with the yield trait. However, high plant height makes sesame susceptible to lodging from high winds, such as a hurricane. The taller the plant, the lower the lodging tolerance. In order to increase the lodging tolerance and the yield, reducing the plant height is a feasible solution for sesame breeders. On the other hand, plant height also affects mechanization production in sesame. Too high plant is not adaptable for mechanical harvest (Langham 2014). As the reel of the header pushes the plants ahead of the combine rather than into the header, the seeds will shake to the ground and may result in high loss (Fig. 2.9a). With Sesaco 70, the reel pulls the sesame into the header (Fig. 2.9b). In some cases, tall plants cross over the auger and do not feed into the combine. Zanten (2001) recommended that the optimal height of sesame for mechanized harvest was 50–150 cm at the Final FAO/IAEA Coordination Research Meeting in 1998. Langham and Wiemers (2002) suggested that the maximum height acceptable for all harvest implements is 150 cm. Lower plants are preferable. Bennett MR believes that the ideal type plant

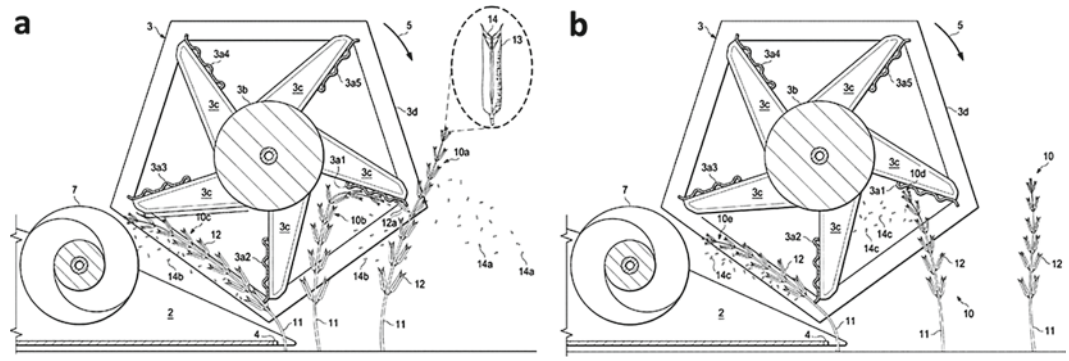


Fig. 2.9 Schematic diagram of mechanical harvest of sesame with different height types. **a** Tall plants harvest diagram with a platform header. **b** Short plants harvest diagram with a platform header (Figures supplied by Ray Langham)

height of sesame should be 120–130 cm tall with 82 cm of reproductive stem with 28 node pairs of capsules (personal communication, 2016).

2.2.2.2 Height to the First Capsule

The height of the first capsule node is measured from the first leaf node or the ground to the bottom of the lowest capsule on the main stem. Some researchers measure from the ground to the lowest capsule. Of the 763 sesame germplasm accessions investigated in 2017, the height of the first capsule varied from 15.6 to 160.0 cm (Fig. 2.5b). Most varieties have medium height of 30–60 cm. Of the 11 varieties with long height to the first capsule ($\geq$ 100 cm), seven are from Africa or Latin America and are day-length sensitive. For example, as cultivated in early May at the Pingyu site, a Tanzanian accession (Tan-south 1) and a Mexican accession (Yori 77-1) produced only a few capsules with shriveled seeds in early September (data not shown). Langham (2017b) reported the international current height to the first capsule range is 9–160 cm. The genotypes with very high first capsules are generally short-day genotypes planted in a long-day environment. Conversely, the genotypes with very low first capsules are long-day genotypes planted in a short-day environment.

Kobayashi (1986) studied the genetic background of the height to the first capsule in sesame and found that the high height to the first flower or capsule is dominant over the low height. The height to the first capsule is affected by both genotype and environment (such as moisture, fertility, time of planting, light quality, and population).

In addition, the height to the first capsule is an important trait for mechanical harvest in sesame production. The cutter bar position should be below the lowest capsule. When the height is too low, the combine will face the risk of picking up stones and damaging the combine. When the height to the first capsule is too high, the plant is generally taller resulting in lower yields. According to the expert suggestions, the recommended height of the lowest capsules should be 15–20 cm above ground for hand-harvested crops, or 15–40 cm above the ground for mechanical harvesting, depending on the machinery and topography (Zanten 2001; Langham and Wiemers 2002).

2.2.2.3 Capsule Zone Length

Capsule zone length extends from the base of the lowest capsule to the tip of the highest capsule with viable seed. It is basically the plant height minus the height to the first capsule. This trait is a yield component trait in sesame.

As shown in Fig. 2.5c, of 763 sesame germplasm accessions, the capsule zone length varied from 13.5 to 129.13 cm. Over 50% of the varieties had the capsule zone length of 70–90 cm. Langham (2017b) reported the international capsule zone range is 12 to 188 cm. The capsule zone length can only be compared within the

same phenotype. Branched phenotypes generally have shorter capsule zone lengths than uniculm phenotypes. However, branched phenotypes with a shorter capsule zone may produce a comparable number of capsules as the uniculm with longer capsule zone.

Most sesame varieties are indeterminate. At the end of the reproductive phase, they may have flowers that do not produce capsules. The non-productive tip length of 763 varieties ranged from 0.0 to 26.5 cm (Fig. 2.5d).

2.2.2.4 Branching Style

Branching style indicates the potential amount of true branching in a line. Langham (2019) used five ratings of the branching style: 0 = Segregating or mixed; 1 = Uniculm—no branching (Fig. 2.10a); 2 = Uniculm except 'weak branches' in open (Fig. 2.10b); 3 = Few branches (3–4 branches) (Fig. 2.10c); and 4 = Many branches (more than 4 branches) (Fig. 2.10d). For sesame, branching generally takes place in nodes on the bottom of the stem below the capsule zone (Langham 2008). The lines with 'many branches' have more nodes below the first capsule, and may have fewer nodes on the main stem. In germplasm libraries of the cultivated sesame, some branched varieties form the secondary branches on branches (Fig. 2.10e), and rarely, there are tertiary and quaternary branches. For the determinate mutant (*dt* 2 type) created by Israeli scientist Ashri A. (Ashri 1998), the tertiary branches form on branches (Fig. 2.10f) (Zhang et al. 2016).

Branching is controlled by a single gene where branched is dominant over non-branched or uniculm phenotype (Nohara 1933; Brar and Ahuja 1979; Kobayashi 1986; Mei et al. 2017). However, the environment also affects the branching style of a sesame variety in the field. In non-thinned populations, it is sometimes difficult to select for both pure branching and uniculm genotypes. When the population is too low, the uniculm plants always produce branches. On the contrary, when the population is too high and the stems cannot get enough sunlight, the branching varieties may not have branches. In a high population, a 'many' branching line may

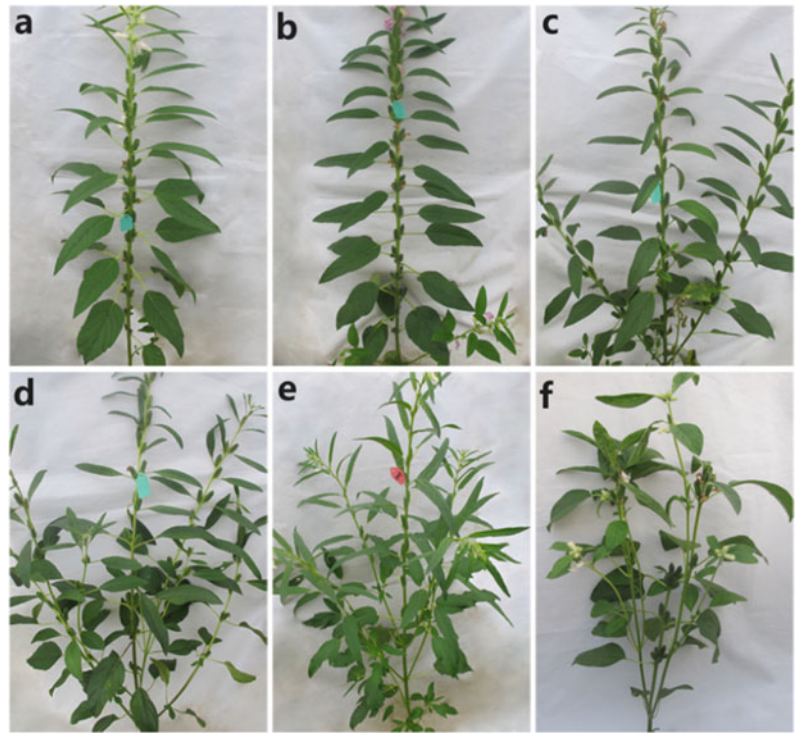


Fig. 2.10 Various branching styles of sesame plantlets. **a** Uniculm. **b** Uniculm with weak branches. **c** Few branches. **d** Many branches. **e** Secondary branches formed on branches. **f** Tertiary branches on branches. (Photographs supplied by Hongmei Miao)

have 0–4 branches. There are some uniculm lines with no branches under all circumstances. Ratings should be made on based several populations to include the end plants of a row. Most uniculm and branching lines have the potential to branch in every leaf axil.

During the Final FAO/IAEA Co-ordination Research Meeting held in 1998, the worldwide sesame breeders proposed uniculm sesame varieties for dense stands under high input conditions and appressed medium branching varieties for low input conditions (Zanten 2001). Branching varieties were recommended to solve the cutting problems of the thick stems of uniculm lines (Beech and Imrie 2001). Branching varieties use the extra space, form more capsules, and also shade the ground to prevent weeds. Generally, the branching reduces plant height but increases the exposure to winds which may lead to lodging. Branching is important for mechanization of sesame cultivation. For the binder, branches help in swathing and combining. In both auger and belt swathers, branched plants will become intertwined and make it easier to move the plants to the windrow. For direct harvest using combines, intertwining helps the header auger move the plants to the feeder housing. However, too much branching makes it difficult to separate the crop at the edge of the header and may result in loss of capsules and/or shattering of seeds (Langham 2007).

2.2.2.5 Height to First Branch

The height of first branch refers to the height from the ground to the base of the first true branch with viable capsules (Langham 2007). For branched varieties, the true branches form below the first flower node pair. There are many phenotypes where the first flowers do not make a capsule, and thus the height to the first flower may be different than the height to the first capsule. Weak branches normally start from the lowest node pair in uniculm line, when the population is low or with ample fertility and moisture. Langham (2017b) showed that the international current known height to the first branch is 1–135 cm. The height to the first branch is highly dependent on the inter-row and intra-row spacing. Higher populations within the row, as well as narrower row width have fewer branches. In high-density populations, the internode length increases, as the plants compete for light. The key is that most leaf axils have the potential to produce a branch, but abundant sunlight should be supplied to that leaf axil.

2.2.3 Leaf

The leaf is an important organ that captures light energy and synthesizes organic compounds through photosynthesis and finally affects the seed yield in sesame (Langham and Wiemers 2002). Recently Langham (2018b) described the traits of tens of botanical traits related to sesame leaf, such as leaf shape, leaf length, leaf blade length, petiole length, petiole groove, ratio of leaf blade length and width, leaf area, basal leaf margin, leaf base shape, leaf apice shape, leaf thickness, leaf venation, leaf enations, and mottled leaf. The results supply the guidance for breeding and botanical research in sesame.

Fig. 2.11 Leaf shape in sesame. Leaf shape of sesame varies from lobed to entire blade, from applanate to curly, from linear to cordate, and from entire to serrated leaf edge. The top group is typical at the base of the plant, the middle group at the middle of the plant, and the bottom group at the top of the plant. (Photographs supplied by Ming Ju)

Sesame is a eudicot with a pair of cotyledon in seed. The leaf organ comprises of leaf blade and petiole. Leaf shape varies and is related with the leaf position and genotypes. In sesame germplasm, leaf shape varies from linear, lanceolate, elliptic, ovate, narrowly cordate, and other specific types (Fig. 2.11). For basal leaves, the extent of the lobe incision is also different (the top row in Fig. 2.11).

Langham (2018b) states that the lower leaves can be entire, lobed, cleft, parted, or divided. Generally, the first leaves increase in size with the 5th or 6th node pair having the largest leaves with various incision depths; the leaf size then decreases up the plant as shown in Fig. 2.12 with measurements at the 5th, 10th, and 15th leaf pair.

The leaf length varies from 4.4 to 42.5 cm; the leaf blade length varies from 4.2 to 25.5 cm; the leaf blade width varies from 0.7 to 33.6 cm; and the petiole length varies from 0.2 to 17.0 cm. Moreover, the pubescence density, leaf reticulate pattern, and leaf color are diverse.

In sesame, the type of leaf arrangement includes opposite, alternate, ternate, and mixed types in the upper half of the main stem at beginning of flowering (Anon 2004). The diversity of leaf shape and the leaf arrangement from up to bottom favors for constructing the ideal plant type and utilizing the ample sunlight. Baydar and Turgut (2000) reported that the lobed leaf shape was dominant over simple or ovate leaf shape. The results obtained from the F_2 generation indicated that inheritance of this trait was monogenic with 3:1 segregation ratio.

Histological sections of the leaves of sesame seedlings results show that leaf blade tissue comprises of upper epidermis (UE), adaxial parenchyma cell (PC (ad)), vascular bundle cell layer (VB), abaxial parenchyma cell (PC (ab)), and lower epidermis (LE). Vascular bundle cells in petiole and primary leaf vein can perform limited secondary growth (Fig. 2.13). In upper and lower epidermis layers, there are stomata. The stomata belong to the buttercup type and

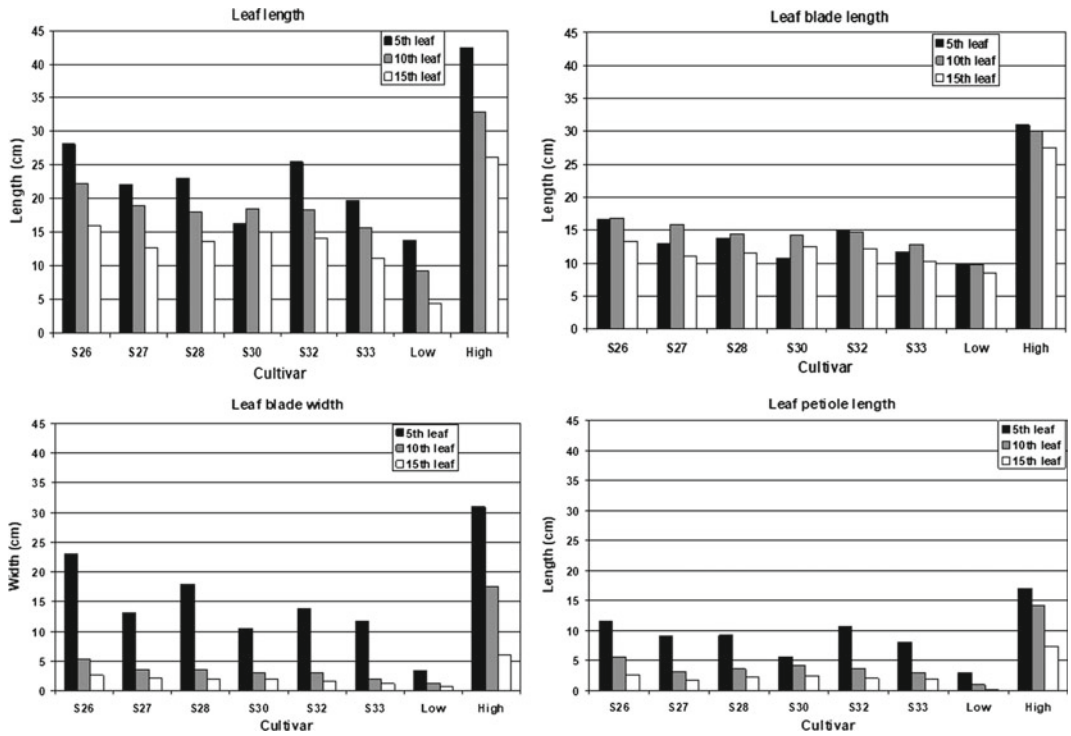


Fig. 2.12 Leaf sizes at various positions of the plant. As shown on six varieties plus the lowest and highest other phenotypes, when measuring the 5th, 10th, and 15th leaf pair from the base, all measurements decrease. The width and petiole length decrease more dramatically than the leaf blade

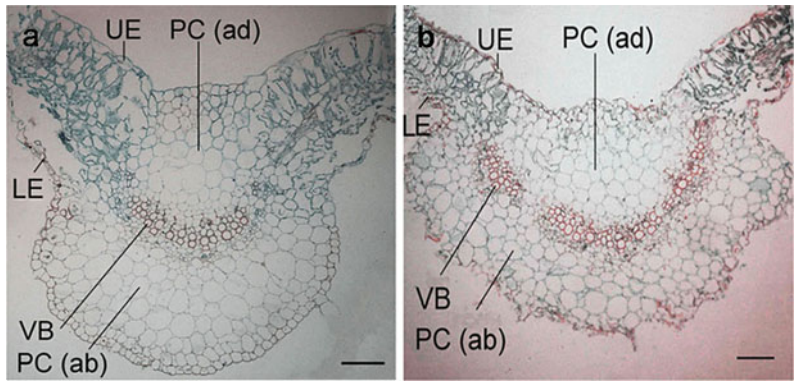


Fig. 2.13 Cross-section of leaf blades of sesame. **a** Cross-section of the seedling leaf blade at the 1 leaf stage of Yuzhi 11. **b** Cross-section of the seedling leaf blade at the 2 leaf stage of Yuzhi 11. The PC (ad) and PC (ad) tissues in veins are horizontally arranged, and the adaxial/abaxial trait is reduced. **UE**: upper epidermis. **LE**: lower epidermis. **PC (ad)**: adaxial parenchyma cell. **PC (ab)**: abaxial parenchyma cell. **VB**: vascular cylinder cells. Bar = 100 μm. (Photographs supplied by Hongmei Miao)

represent the characteristics of family Pedaliaceae (Li 1987).

The surface structure of leaves is related with the drought tolerance in sesame. Scanning electron microscope (SEM) shows there are four types of leaf hairs on the leaf surface of sesame seedlings: non-glandular hairs, long stalk glandular hairs, short stalk glandular hairs, and mucilage hairs (Fig. 2.14). The secretions of glandular hairs contain a variety of compounds. Varieties with different drought tolerance present different glandular trichome structure (Su et al. 2016). Thus, the specific types of glandular trichome structure and the secretion components can be applied for sesame germplasm identification and the drought tolerance evaluation (Su et al. 2016). Langham (1945a) reported that genotypes with many glands had drought tolerance under drought conditions. In periods of excess rainfall, genotypes with fewer glands had more waterlogging tolerance. He found that glands were dominant over glabrous in leaves. Sun et al. (2009) reported differences in leaf secretions in relation to waterlogging.

Although there have been studies on the leaf glands, very little is known whether the pubescence on the other organs are related to drought and waterlogging tolerances. The density, types, and lengths of pubescence vary on the cotyledons, stems, leaves, petioles, outer and inner flower corollas, flower lips, calyxes, and capsules (Fig. 2.15) (Langham 2017b). Several studies have shown that the pubescence presence is independent from one organ to the next (Bisht et al. 1998; Mahajan et al. 2007; Frary et al. 2015).

In sesame, leaf angle and leaf area index are important traits for yield breeding. Most genotypes have acute angle leaves at the top of the plant, but the angle increases the lower the leaf. There are genotypes where the acute angle persists. In general, when the fertility potential and capsule setting is low, the leaf angle is acute (Fig. 2.16a, b). For most varieties with high yield in China, leaves in up stem are more horizontal and are not as acute (Fig. 2.16c, d). As to a specific sesame line, improving the plant type with acute leaf angle will help the leaves receive and utilize sunlight more efficiently, resulting in higher yield.

The largest leaves in terms of area form between the 3rd and 6th node (Langham 2018b). During this stage, sesame plants are transitioning from the vegetative phase into the reproductive phase, and more nutrition and energy are needed. Babu et al. (2004) estimated genetic parameters in 18 white seeded genotypes of sesame and reported that additive and non-additive gene action appeared to be important for the leaf area index trait. Saravanan and Nadarajan (2003) reported that non-additive variance played a major role for leaf area index.

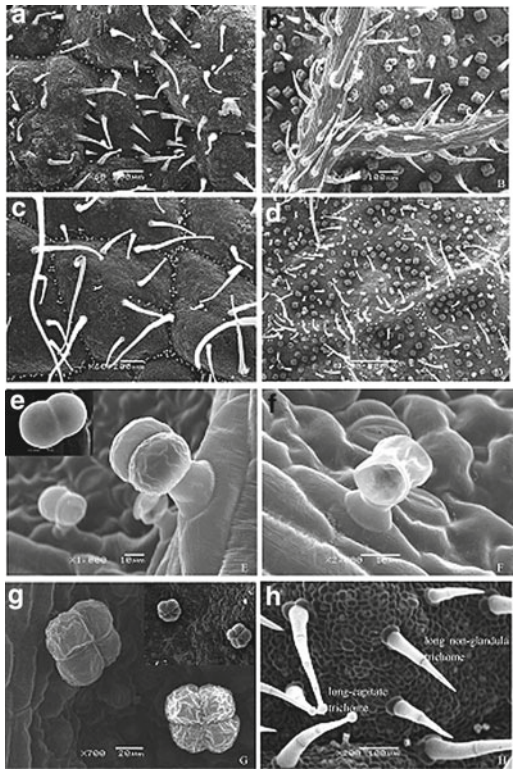


Fig. 2.14 Normal glandular trichomes on the surface of sesame leaves by scanning electron microscope (SEM). **a** and **c** The glandular hairs on adaxial (dorsal) surface of Rongxian black sesame and Yuzhi 11, respectively. **b** and **d** The glandular hairs on abaxial (ventral) surface of Rongxian black sesame and Yuzhi 11, respectively. **e** The short stalk capitate trichomes type 1. **f** The short stalk glandular trichomes with two cell head in the post secretory phase (PPS). **g** The peltate trichome typela (with four cell head). **h** The long-capitate trichomes and type I. (Cited from Su et al. 2016)

As to the leaf edge of sesame, several mutants induced naturally or through EMS mutagenesis resulted in different leaf shapes (Fig. 2.17). Figure 2.17a shows a normal leaf. A natural mutant *cl*1 displays the up curly leaf shape trait during the entire development life (Fig. 2.17b). Some enations are present on the veins of the back blade of the leaves in mutant *cl*1 (Langham 1946, 2001; Langham and Rodriguez 1946). Besides the mutation of leaf edge and the enations, there may be enations on the flower corollas; bent styles in the flowers resulting in a loss of fertility; a twisted stem; and extra layers of cells across the carpels resulting in an indehiscent capsule (closed capsule), when dry. Cross-section observations of leaf blades indicate that the lower epidermis in *cl*1 is not as smooth as that in the wild type. The abaxial parenchyma cell (PC (ad)) and adaxial parenchyma cell (PC (ad)) tissue end to be horizontally arranged in leaf veins (unpublished data, Haiyang Zhang). Another curly leaf mutant *cl*2 presents curly leaf with curly stem (Fig. 2.17c). The curly trait is controlled by a semi-dominant gene. The homozygous dominant line presents the curly leaf and twisted stem, capsule number in the fragile stem reduces obviously. For the heterozygous line, the stem is normal and erect, only leaves are curly. Other mutants related with leaf color and leaf shape are reported in Chap. 5 of this book. There are genotypes that under low moisture stresses, the leaves facing east toward the sun will curl up, while the leaves on the west side of the plant are

Fig. 2.15 Differences in the pubescence in the capsules. **a** Glabrous pubescence. **b** Medium pubescence. **c** Profuse pubescence

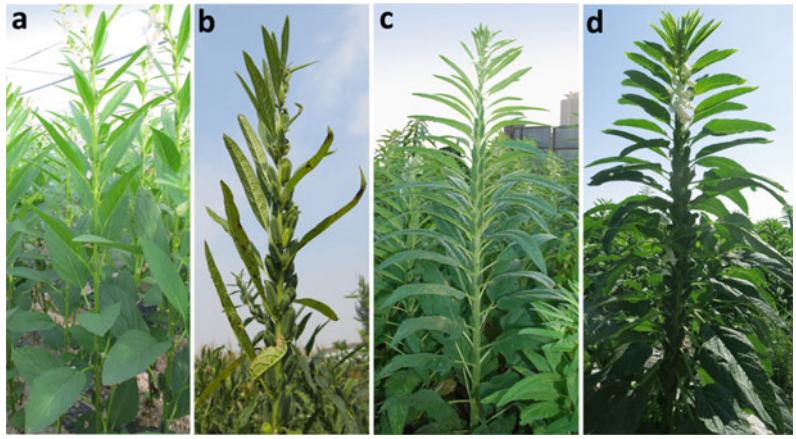


Fig. 2.16 Various leaf angle types of sesame. **a** Sharp angle of 1 capsule per leaf variety. **b** Sharp angle of 3 capsule per leaf variety. **c** Gentle acute angle of sesame variety. **d** Applanate angle of short internode length variety. (Photographs supplied by Hongmei Miao)

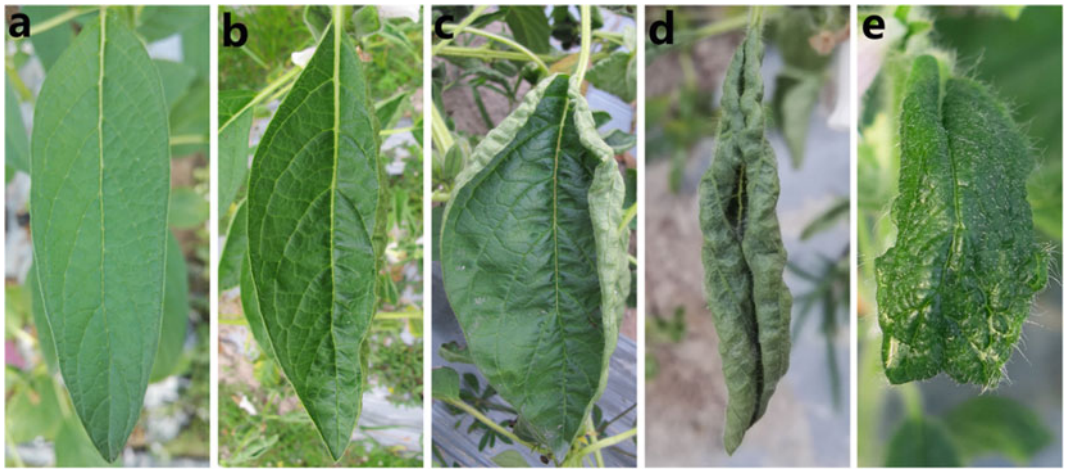


Fig. 2.17 Leaf shape with various curling types of sesame. **a** Normal leaf. **b** Gentle curly leaf with enations at the veins of the *cl*1 mutant found by DG Langham. **c** Curly and shrinking leaf (mutant *cl2*) **d** Whole curly leaf. **e** Under-curly leaf with shrinking phenomenon, named by DG Langham (1945b) as wrinkled leaf and controlled by a recessive allele. (Photographs supplied by Hongmei Miao)

normal. As the sun moves to the west in the afternoon, the leaves on the west side will curl while the leaves on the east side return to normal (data not shown).

2.2.4 Flower

As the plant enters from the vegetative phase into the reproductive phase, buds will form. The sesame flower is bisexual. The inflorescence initiation and differentiation directly affect the yield per plant. This section presents the morphology of sesame flower, inflorescence differentiation, and spontaneous outcrossing.

Sesame flower grows in the leaf axil on short pedicel. The sesame flower consists of pedicel, calyx, corolla, stamen, and pistil. The stamen consists of filament and anther, while the pistil consists of ovary, style, and stigma. The flower corolla is tubular and campanulate with long lip (Fig. 2.18). Prior to opening, the flower buds are

yellow-green. When the flower opens (about the same time fertilization takes place), the color of flower corolla varies from white, light pink, pink, red, violet, to purple (Fig. 2.18). The purple flower is dominant over white flower (Nohara 1933; John 1934; Pal 1934; Sikka and Gupta 1949; Khidir 1973; Kobayashi 1986). There are abnormal flowers that are green, and the lip does not open, and yet there is fertilization, and a capsule forms. The flower size also varies in sesame. In *Sesamum*, to our knowledge, the largest flower is from the wild species *S. angolense* with the corolla length of 6.8 cm. The longest known corolla length in sesame is 4.5 cm (Langham 2018c). Besides the corolla, other 13 flower-related traits including the calyx tip color, calyx lobe length, calyx hairiness, calyx hair, corolla hairiness, interior corolla color, corolla interior pigmentation, lower lip color, the longest lip length, the absence/presence of foveola,

Fig. 2.18 Morphological traits of sesame flower. **a-p** Flower type variation in corolla color, corolla size, petal shape, lip color, and shape. **e**, **g**, **m**, **n**, **o**, and **p** Indicate the flowers of the wild *Sesamum* species or interspecies progeny. (Photographs provided by Hongmei Miao)

anther filament color, anther connective tip gland, and anther style length are used to describe the morphological traits of the sesame flower (Anon 2004). Langham (2018c) provides additional 45 morphological traits with flower.

Langham (1947a) reported a wide variation of coloring inside the corolla where the anthocyanins (purple) and flavonols (yellow) have different patterns (smear, flake, or dots) in different parts of the lip in relation to the foveola in different intensities (from absent to strong). Langham (1947b) reported 16 different morphological traits within the flower.

Genotypes with a single flower per axil form two extra nectaries in the leaf axil on both sides of the flower. Extrafloral nectaries are believed to be rudimentary flowers. In different sesame germplasm accessions, the color, size, and number of nectaries per axil are variable (Fig. 2.19). In Chinese germplasm accessions, co-existence of auxiliary two nectaries and three capsules per axil has been found (Fig. 2.19d). These triple capsule lines have the potential to develop 5 or even 7 flowers per leaf axil. The nectaries have nectar in the center and attract insects and hummingbirds to feed. For the wild species, the nectary usually is purple and small. For mechanized sesame production, the nectaries may dry down to a size and weight similar to the seed, and they are difficult to clean out of the seed.

In genotypes with three flowers per leaf axil, initially there is one flower per leaf axil. Then, 2–5 days later, two axillary flowers may appear besides capsule forming from the first flower (Langham 2018c). There are genotypes with a propensity to produce 5 and even 7 flowers in low populations in strong sunlight. Compared to the axillary flowers, the central flowers produce longer capsules with more seeds per capsule. Under low fertility and/or moisture, high temperature, and/or moisture conditions, the development of the axillary flowers (sometimes even central flowers) may result in the abscission of flowers and capsules.

2.2.4.1 Inflorescence Differentiation

In sesame, the flower bud initiates from the continuous differentiation of the apical meristem throughout the whole reproductive phase. Thus, sesame is considered an indeterminate crop with over a month in the reproductive phase. However, there are a few determinate varieties, which can form a terminal flower cluster, as the shoot apical meristem is converted into a terminal floral meristem (Melzer et al. 2008; Hanano and Goto 2011). For sesame, there are two types of determinate mutants. One is the *dt*1 type mutant DS899. This determinate mutant was created from sibling line of Yuzhi 11 (applied for the Sesame Genome Project) using ethyl methane sulfonate (EMS) mutagenesis in 2009 (China patent no. ZL2015108760163). The *dt*1 mutant line can develop 8–20 capsule node pairs before the development of the inflorescences terminates. The other is *dt*2 type determinate mutant which is

Fig. 2.19 Sesame nectary number and type in leaf axil. **a** Capsules without nectaries. **b** Small nectaries. **c** Large nectaries. **d** Nectary accompanied on plant with three capsules per leaf axil. **e** Purple nectaries of wild species. (Photographs provided by Hongmei Miao)

induced by gamma ray (5000 Grey) (Ashri 1998). Differing from the wild type and *dt*1 type, *dt*2 mutant has a limited number (3–5) capsule node pairs on the main stem but will develop branches with additional capsules. This number of node pairs on the main stem is not affected by environmental factors (Zhang et al. 2016). Gene function analysis proves that the determinate traits are controlled by the alleles of the *SiDt* gene. More information about gene cloning and its function will be introduced in the following chapters.

Scanning electron microscope (SEM) analysis shows the initiation and differentiation of the inflorescences in sesame (Fig. 2.20) (unpublished data, Haiyang Zhang). The indeterminate wild type (var. Yuzhi11), *dt*1 line (var. Yuzhi DS899) and two *dt*2 derived lines (08TP092 with 4-locule capsule trait and US92 with 6-locule capsule trait) are analyzed. For the wild type and *dt*1 types, the apical meristem at the stem tip is leaf primordial at seedling stage. With the continuous differentiation, enlargement, and bulge, new apical leaf primordial and lateral flower meristems form (Fig. 2.20a, b). Flower primordial meristems are laterally arranged surrounding the apical meristem (Fig. 2.20e). Finally, the apical meristem of *dt*1 becomes a terminal flower meristem (Zhang et al. 2016). For *dt*2, the apical meristem forms flower meristem mat 2 leaf pair old seedlings (Fig. 2.20c). Terminating the inflorescence differentiation occurs earlier, compared with the wild type and *dt*1 type (Zhang et al. 2016). During the apical meristem development, leaf primordial develops into bracts. The inflorescence meristem of sesame divides into calyx primordial, petal primordial, stamen primordial and pistil (Fig. 2.20e, f). For the 4-locus-capsule genotypes (*Dt, dt*1, and *dt*2), there exist four petal primordial and four stamen primordial in the inflorescence meristem. Meanwhile, the 6-locule-capsule genotype (*dt*2) presents six petal primordial and six stamen primordial (Fig. 2.20 g). The results indicate that the differentiation of petal and stamen number is diverse in various germplasm accessions.

2.2.4.2 Anther and Pollen Development

The sesame anthers are white and full of pollens (Fig. 2.21a). After dying using the Alexander staining method, the pollen grains are red and ball-like (Fig. 2.21b). According to the morphologic characteristics of its microspores, Yang et al. (2008) divided the sesame pollen development process into seven stages, i.e., microsporocyte formation stage (bud length <2 mm), meiosis stage (2 mm $\leq$ bud length < 3 mm), tetrad stage (3 mm $\leq$ bud length < 4 mm), early microspore stage (4 mm $\leq$ bud length < 5 mm), middle microspore stage (5 mm $\leq$ bud length < 6 mm), late microspore stage (6 mm $\leq$ bud length < 7 mm), and pollen maturation stage (bud length $\geq$ 7 mm).

Transmission electron microscopy (TEM) and SEM images show that the male gametes development initiates with the differentiation of stamen primordia and comprises of seven stages in sesame (Fig. 2.22; Yang et al. 2008; Guo et al. 2014). In microsporocyte formation stage, the stamen primordia develop into microsporocytes and four layers of anther wall (Fig. 2.22a, b). The microsporocytes are large and specific. During performing meiosis, the metabolism activities of microsporocytes are increasing (Fig. 2.22c). Lipid globules appear, and tapetal cells walls become crooked. Probacula form between the callose and plasma membrane. Then the tetrad cells form and are surrounded by thick callose wall (Fig. 2.22d). The tapetal cells are dissolved. In middle and late microspore stage, the microspores are nearly spherical with the nucleus locating in the center soon (Fig. 2.22e, f). Tapetal cells disintegrate and many ubisch bodies are secreted. Subsequently, the tapetal cytoplasm density decreases (Fig. 2.22g, h). With the uneven mitosis, the microspores form and develop into mature pollen grains (Fig. 2.22i). Meanwhile, the sterile pollen grains develop abnormally with irregular shape. Compared with the fertile pollen, the sterile pollens in segregated lines of ms86-1 are aborted as a disc. Semi-fertile lines form aborted and partial aborted pollens (Fig. 2.22j–l).

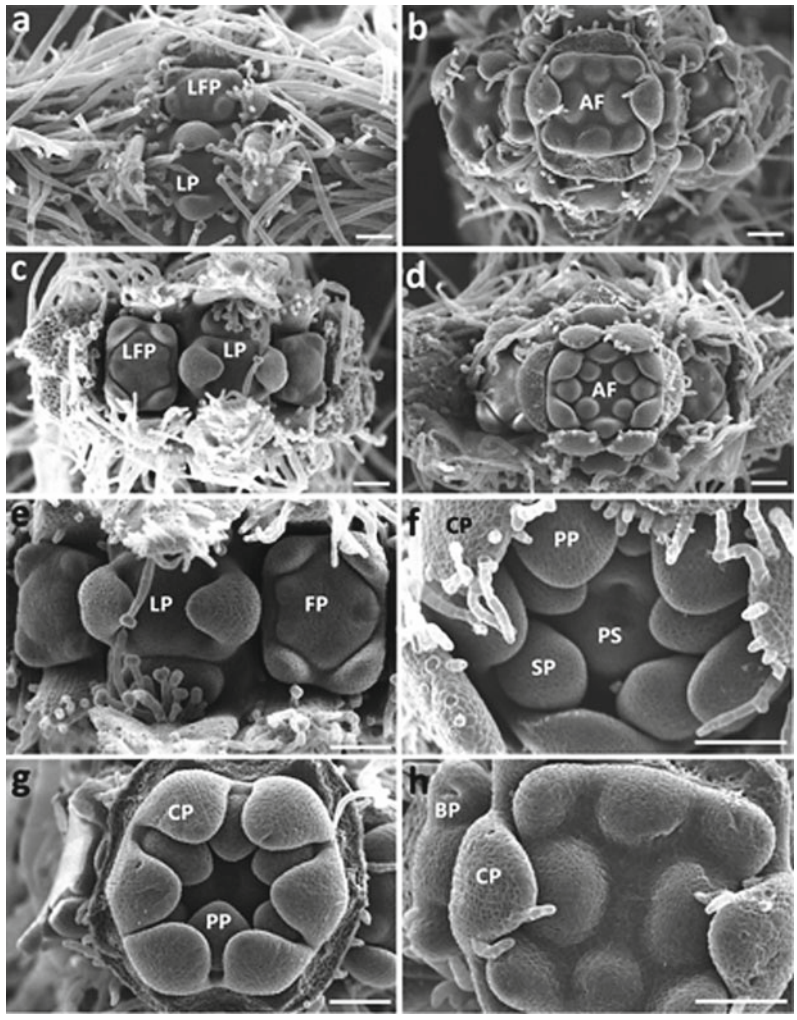


Fig. 2.20 Inflorescence initiation of determinate sesame mutants and the wild type by scanning electron microscope. **a** Inflorescence initiation of Yuzhi 11 (indeterminate, wild type) at 2 leaf pairs (×100). **b** Inflorescence initiation of Yuzhi DS899 (determinate, *dt*1 type) at 2 leaf pairs (×100). **c** Inflorescence initiation of US92 (determinate, *dt*2 type) at 2 leaf pairs (×100). **d** Inflorescence initiation of 08TP092 (determinate, *dt*2 type) at 5 leaf pairs (×100). **e** Differentiated apical leaf primordial and lateral flower primordial of Yuzhi 11 (×200). **f** Flower meristem differentiation of 4 locule capsule type (var. Yuzhi 11) (×300). **g** Flower meristem differentiation of 6 locule capsule type (var. US92) (×300). **h** Flower meristem differentiation of 4 locule capsule type (var. 08TP092) (×200). **AF**: apical flower meristem. **AFP**: apical flower primordia. **BP**: bract primordial. **CP**: calyx primordial. **LFP**: Lateral flower primordial. **LP**: Leaf primordia. **PP**: Petal primordial. **PS**: Pistil stigma. **SP**: Stamen primordial. Bar = 100 μm. (Photographs provided by Hongmei Miao)

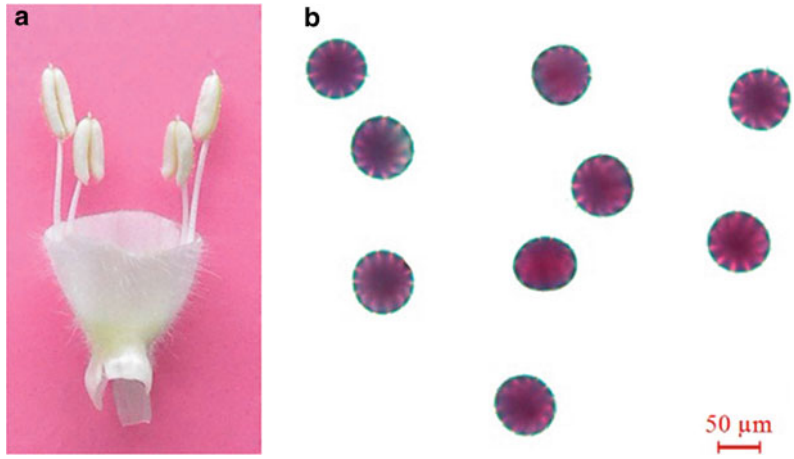


Fig. 2.21 Morphology of sesame mature anther and pollen. **a** Mature anther. **b** Mature and fertile pollen dyed using Alexander staining method. (Photographs provided by Hongmei Miao)

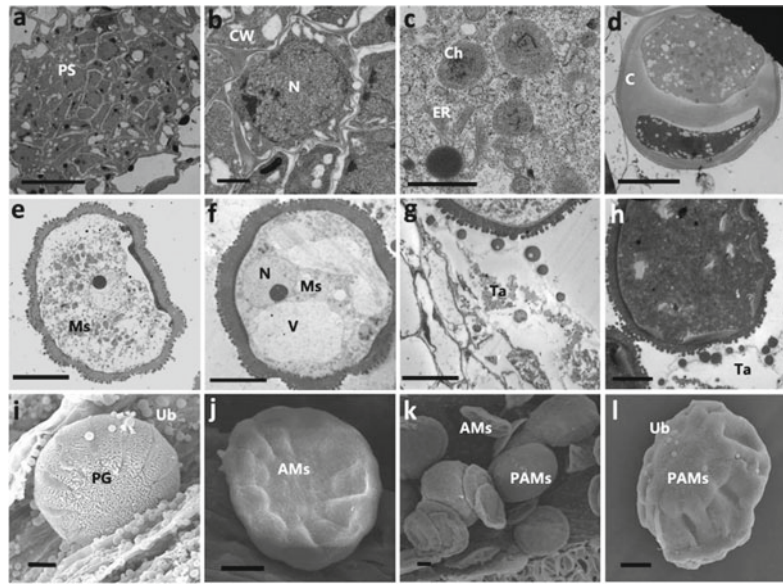


Fig. 2.22 Ultrastructure of microspore formation in sesame. **a** Microsporocyte formation stage. **b** Microsporocyte. **c** Meiosis stage. **d** Tetrad stage. **e** Early microspore stage. **f** Middle microspore stage. Spherical microspore presented with several large vacuoles. **g** Tapetal cells with sparse cytoplasm in late microspore stage. **h** Mature pollen with several apertures on the surface with ubisch bodies. **i** Pollen maturation stage. **j** Aborted microspore in genetic male sterile (GMS) line ms86-1. **k:** Aborted and partial aborted microspores in semi-fertile line of ms86-1. **l** Partial aborted microspore with some ubisch bodies. **PS**: Pollen sac. **CW**: cell wall. **N**: nucleus. **Ch**: chloroplast. **ER**: endoplasmic reticulum. **C**: Callus. **Ms**: microspore. **V**: vacuole. **Ta**: Tapetum. **Ub**: ubisch bodies. **PG**: Pollen granule. **AM**: Aborted microspore. **PAM**: Partial aborted microspore. In photos b and c, Bar = 1 μm. In the other 10 photos, bar = 10 μm. The fertile, sterile, and semi-sterile plants of ms86-1 are applied for analysis. (Modified from Yang et al. 2008; Guo et al. 2014)

2.2.4.3 Cross-Pollination Frequency

Sesame is considered as a self-pollinated crop. However, the cross pollination by insects is common with the current known range of 0–68% (Howard et al. 1919; Mohammad and Alam 1933; Richharia and Persai 1940; Sikka and Gupta 1949; Kumar and Gokhale 1952; Muhammed et al. 1965; Khidir 1972, 1981; Yermanos 1980; Pathirana 1994). Most of the above studies used a similar methodology to Langham (1944). According to the investigation results of cross-pollination of non-glabrous seedlings in progeny of open-pollinated glabrous lines, the natural cross pollination was 4.6%. Van Rheenen (1980) compared three cultivars and found that in one cultivar the cross-pollination was 2.7%, while the second was 51.7%, illustrating that different cultivars had different outcrossing rates. The results reflected that controlled self-pollination is advisable for the maintenance of germplasm collections. Yermanos (1980) found that the cross pollination reached to less than 1%, when the sesame was surrounded by cotton and other crops. In Moreno, CA, cross pollination frequency was 68% in a field where the sesame was the only blooming plant in a semi-arid area. Sun et al. (2015) investigated the cross pollination rate using two groups of sesame varieties under mixed planting pattern during 2010–2012. One is the determinate/indeterminate cultivar group. Of which, the determinate cultivar is *dt*2 type closely linked with capsule indehiscence trait (Ashri 1998). The other is mono-flower/tri-flower per axil cultivar group. The individual trait of these two groups is controlled by one gene pair. Results showed that the natural crossing pollination frequency varied from 5.13 to 23.35% with a mean value of 12.69%. The value can be affected by planting region, year, and cultivar genotype. Thus the suggested isolation distance is above 400 m with a buffering zone of 8–12 m, in order to guarantee the seed purity of above 99.9%. Anon (1998) specified a buffer of 10 m around the certified or foundation seed.

Visits by insects have been shown to increase the fertility within the capsules and to increase yields (Panda et al. 1988, 1989; Montilla and Cedeno 1991; Patil et al. 2000; Falusi 2007; Blal et al. 2012, 2013; Mahfouz et al. 2012; Nithya et al. 2012; Pashte et al. 2015). Insects (*Alydidae, Anthophoridae, Apidae, Calliphoridae, Coccinellidae, Coreidae, Crabronidae, Eumenidae, Formicidae, Halictidae, Ichneumonidae, Lycaenidae, Megachilidae, Muscidae, Nitidulidae, Nymphalidae, Pentatomidae, Peridae, Sarcophagidae, Scoliidae, Sphecidae, Syrphidae,* and *Vespidae* families) are the main source of cross pollination, while wind has little effects on crossing pollination in sesame (Sikka and Gupta 1949; Mahfouz et al. 2012; Mahmoud 2012; Nithya et al. 2012; Kamel et al. 2013; Pashte and Shylesha 2013a, 2013b; Said et al. 2013; Ngongolo et al. 2015; Sajjanar and Eswarappa 2015; Sun et al. 2015; Fohouo and Nepide 2018).

2.2.4.4 Day Length Sensitivity

Sesame has both short-day and long-day genotypes. If a short-day genotype is grown under long-day conditions, the late maturity sesame tends to extend the vegetative stage (Rhind 1935). If a long-day genotype is grown under short-day conditions, the plants will flower and terminate very early. Day length insensitive sesame can produce flowers and capsules under short-day and long-day conditions, even though the latter is better. Ghosh (1955) reported that short photoperiod could advance the flowering process particularly in the medium-late varieties. According to the flower stage observation of the 763 worldwide sesame germplasm accessions in China (Pingyu), nine exotic varieties did not flower until the end of August. The flowering stage of the other 754 accessions varied from 18.3 to 38.3 d, with the average flowering days of 30.4 d (unpublished data, Haiyang Zhang). The 4 varieties with short flowering stage ($\leq$ 20 days) are from the north of China such as Jilin and Liaoning provinces and express the early-maturing trait. Of the 13 cultivars with long

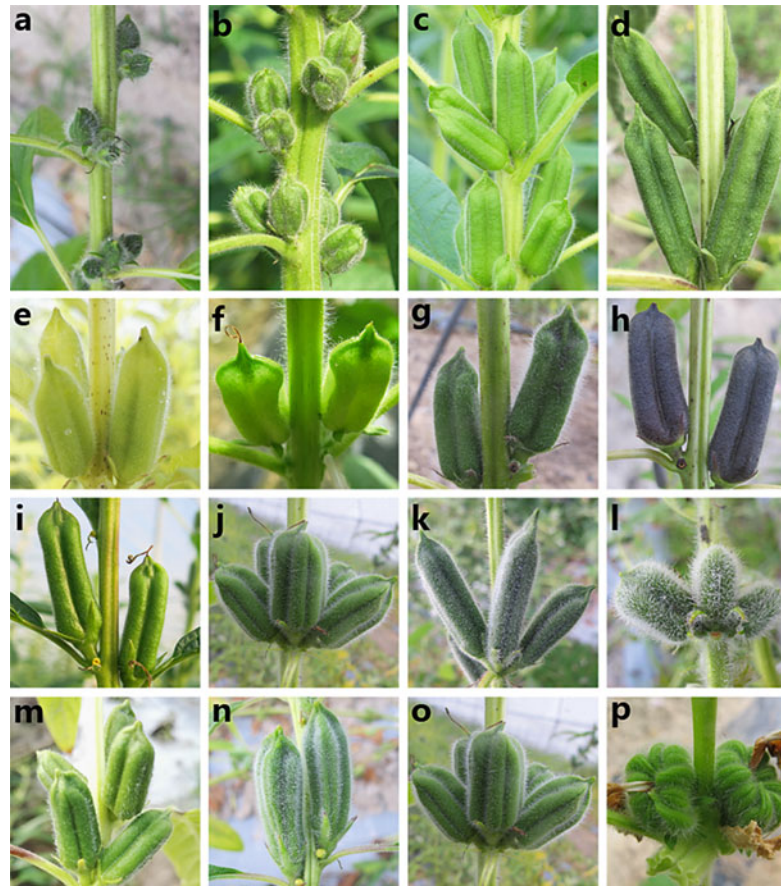


Fig. 2.23 Morphology of sesame capsule. **a** Small-sized capsule. **b** Middle-sized capsule. **c** Large-sized capsule. **d** Long capsule. **e** Yellow capsule. **f** Green capsule. **g** Green capsule with purple stripes. **h** Purple capsule. **i** Glabrous capsule. **j** Capsule with sparse pubescence. **k** Capsule with short pubescence. **l** Capsule with dense and long pubescence. **m** Bicarpellate capsule. **n** Tricarpellate capsule. **o** Quadricarpellate capsule. **p** Multi-carpellate capsule. (Provided by Haiyang Zhang)

flowering stage ($\geq$36 days), 2 varieties were sent from Mozambique and Greece (probably an accession from the tropics) and cultivated several years in China. The other 11 short-day varieties were Chinese domestic varieties from the middle and the south of China. The flowering stage affects the capsule development and the final yield. Short-day varieties grown in a long-day environment and long-day cultivars grown in a short-day environment have poor yields.

2.2.5 Capsule

Sesame capsule is the organ producing seeds. Sesame capsules form from flowers in the leaf axil at about the 4th to 6th node pairs. The morphology of capsule-related traits such as capsule size, capsule color, carpel number, capsule pubescence, and dehiscence varies in sesame germplasm (partially in Fig. 2.23). To reflect the characteristics of sesame capsule, four key traits of capsule, i.e., number of capsules per leaf axil, carpel number, capsule size, and shatter resistance are usually applied for genetics analysis.

2.2.5.1 Capsule Number Per Leaf Axil

Capsule number per leaf axil trait is one of the yield components. Anon (2004) defines the capsule arrangement trait as mono capsular or multi capsular type. Many descriptions classify it as single capsule and 2 or more. Even the genotype is 3 capsules per leaf axil, not all axils have 3 capsules. Others classify the capsule per leaf axil as single (1) and triple (3) capsules (Fig. 2.24a, b). Langham (2017b) divided the capsule number per leaf axil into 5 ratings. i.e., 0 = Segregating or

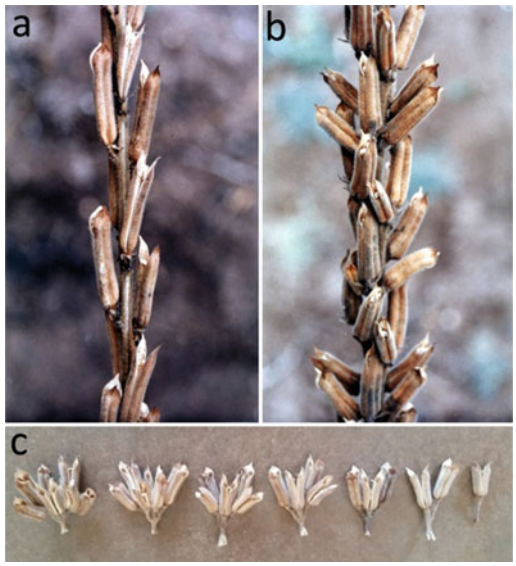


Fig. 2.24 Capsule number per axil. **a** Single capsule per axil. **b** Triple capsules per axil. **c** All these capsules (7, 6, 5, 4, 3, 2, and 1 from left to right) are taken from the same plant. Plant is provided by Wongyai W. (Modified from Langham (2008))

mixed, 1 = Single capsule, 3 = Triple capsule. 5 = Five capsules, and 7 = Seven capsules. Five capsules per leaf axil are common in Chinese germplasm library. In one plant, the number of capsules per leaf axil varied from 1 to 7 (Langham 2017b) (Fig. 2.24c). Weiss (2000) reported an axil forms up to 8 capsules.

The predominant number of capsules per leaf axil is usually expressed in the middle zone of the capsule zone. However, there are cases where a single capsule genotype may have 1–5 node pairs of triple capsules in low populations with high light/moisture/fertility conditions. Conversely, a triple capsule genotype may only have single capsules in high populations with low light/moisture/fertility conditions. Langham (1945a) found that lines with three leaves per node normally lead to three capsules per node.

It is preferable to investigate the capsule traits, after the leaves have fallen off. The capsule number varies in stem part and is affected by genotype and environment. Most triple capsule lines have single capsules at the bottom and the top of the stem.

Rare cultivars consistently form three capsules in every node of the stems (Langham 2008).

Langham (2007) found that the axillary capsules on triple capsule lines formed fewer seeds and lower 1000-seed weight, compared with the central capsules. In a triple capsule line, the central flower always forms 3–5 days earlier than axillary flowers. The seed weight in axillary capsules takes up 69.9–87.3% of that of the central capsule. To date there have been no studies that have determined whether 1 or 3 capsules have higher yield level. Many sesame breeders including Langham DG, Kinman ML, Yermanos DM, Mazzani B, Beech DF, and Bennett MR believe that single capsule lines do better in rainfed or semi-irrigated conditions, while triple capsule lines do better in full irrigation or high and timely rainfall regions (Langham 2007). Several investigations have shown the ratio of capsule formation from a flower varies from 61.4 to 82.8%. Triple capsule lines have the lowest ratios (i.e., 59.2% and 51.6% for single stem and few branch lines, respectively) of capsules versus flowers.

Single capsule is dominant over triple capsule with a single allele (Nohara 1933; John 1934; Pal 1934; Langham 1945b; Sikka and Gupta 1949; Culp 1960a; Baydar and Turgut 2000; Yol and Uzun 2012).

2.2.5.2 Capsule Carpel Number

The number of carpels per capsule is an important trait in sesame (Fig. 2.25). Anon (2004) classified the carpel number into bicarpellate and tetracarpellate types. Langham (2017a) divided the sesame capsule into five groups with various values, i.e., 0 = Segregating or mixed, 2 = Bicarpellate, 3 = Tricarpellate, 4 = Tetracarpellate, and 99 = Other, according to the most common capsule types. Some rare capsules have 6–10 carpels (Yermanos et al. 1964; Fig. 2.25a). The predominant number of carpels per capsule is observed better in the middle half of the capsule zone.

In a bicarpellate capsule, in each carpel there are two locules are separated by false membrane (indicated by red line in Fig. 2.25b) for a total of four locules each with a row of seeds. An

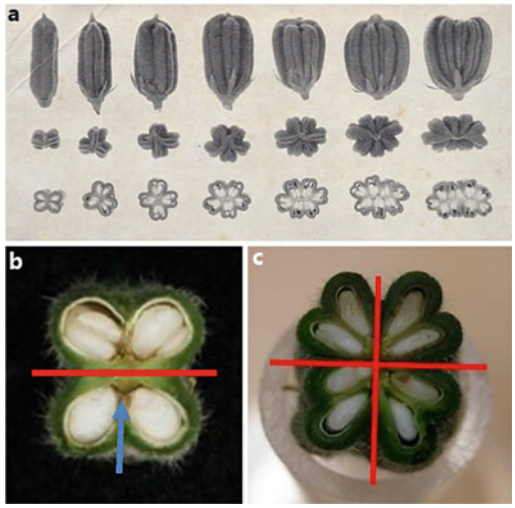


Fig. 2.25 Capsule carpel type in sesame. **a** Carpel number types from left to right are bicarpellate, tricarpellate, quadricarpellate, and rare capsule types with more than one 4+ carpel capsule, respectively (photo cited from Yermanos et al. (1964)). **b** Carpel section (separated by red line) of bicarpellate capsule. **c** Carpel section (separated by red lines) of tetracarpellate capsule. (Photos supplied by Ray Langham)

exception is the seamless capsule which does not form false membranes resulting with two rows of seed in the carpel but no separation of the rows. In a teracarpellate capsule, there are four carpels with eight locules (separated by red lines in Fig. 2.25c). In sesame breeding research, many breeders classify the lines and the capsules based on locule number instead of carpel number. Hildebrandt (1932) felt that sesame should be divided into two subspecies based capsule the number of carpels: bicarpellate and quadricarpellate.

In most cases, bicarpellate is dominant over tetracarpellate with a single allele (Abe 1919; Nohara 1933; John 1934; Langham 1947c; Khidir 1973; Van Rheenen 1981; Kobayashi1986; Gutierrez et al. 1994; Baydar and Turgut 2000; Padmasundari et al. 2010). Field investigations indicated that carpel number of capsule trait is also affected by environment. There are so many examples of plants with mixed number of carpels in sesame. Some bicarpellate plants will have one or more nodes near the center of the capsule zone that have tri- and/or tetracarpellate capsules, and vice versa. Most tri- and tetracarpellate plants will begin and end with bicarpellate nodes (Langham 2017a). In some capsules, there are double rows of seeds within a locule, especially at the bottom of the locule. The seeds seem more constricted with irregular shape (Langham 2007).

Theoretically, more rows of seeds in capsules will result in higher yield. To date, most breeders find that bicarpellate lines outyield tetracarpellate. Van Rheenen (1981) performed a backcrossing of a multi-locular capsule to a four-locular line. However, the yield reduced 34%. Yol and Uzun (2012) postulated that the lack of yield resulted from the inadequacy of the assimilation of photosynthetic production per leaf. The extra capsules and locules in the quadricarpellate types could not be filled adequately.

2.2.5.3 Capsule Size

To describe the capsule size of sesame, the indicators of capsule length, capsule width, and capsule thickness are applied. According to the method of Langham (2017a), capsule length is measured from the bottom to the top of seed chamber outside of the capsule. The length range of capsule is not affected whether calyx and pedicel exist or not. The tip of the capsule is not included (Fig. 2.26a). Langham (2017a) reported the international current capsule length range is 1.3–7.0 cm. Meanwhile, the capsule width refers to the width from the edge of one carpel to the edge of the other carpel (Fig. 2.26b). Langham (2017a) reported the international current known range of capsule widths is 0.5–2.3 cm. For capsules with more than two carpels, the longest distance of carpels should be measured (Fig. 2.26b). As to capsule thickness, the measured side of bicarpellate capsules refers to the narrowest side where the capsule will split at maturity. For capsules with more than two carpels, the thickness indicates the longest distance within a carpel (Fig. 2.26c, d). Langham (2017a) reported the international current known range of capsule thicknesses is 0.5–0.9 cm. During

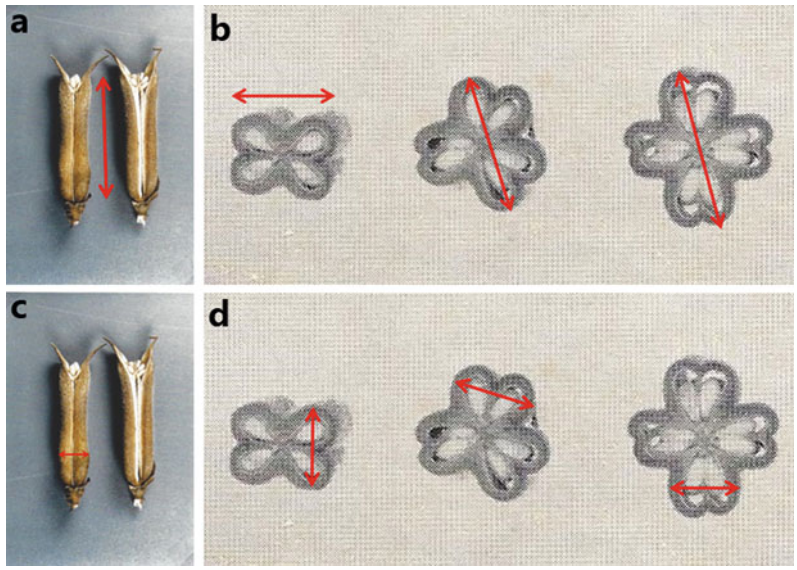


Fig. 2.26 Measurement method of capsule size of sesame. **a** The capsule length measured from the bottom to the top of seed chamber outside (indicated by double head arrow). **b** The width of capsule is measured from the edge a carpel to the edge of the other carpel (indicated by double head arrow). **c** and **d** The thickness of capsule is measured across a carpel (indicated by double head arrow). (Photos provided by Ray Langham)

measuring, fully elongated capsules (central capsules on a triple capsule line) are chosen from the middle of the capsule zone from 5 representative plants.

In China, the capsule size is one of the considerable yield components traits. Chinese scientists usually investigate the capsule traits with the green mature capsules before drying down or split. The capsule length indicates the measured distance from the tip to the bottom of a capsule. Of the 763 germplasm planted in triple replications in 2017, the ranges of capsule length, width, and thickness varied from 1.87 to 4.98 cm, 0.61 to 2.20 cm, and 0.57 to 1.46 cm, respectively (Fig. 2.27a–c). Meanwhile, the seed number per capsule varied from 17.4 to 117.7 (Fig. 2.27d). 450 (59.0%) accessions had 50–70 seeds per capsule (Haiyang Zhang, unpublished data).

Sesame is a productive crop, as the seed number per capsule and per plant is very high. Increasing capsule length indeed can increase the volume of the seed chamber. Wongyai and Juttpornpong (1992) found that capsule length and width are correlated with seed weight. The capsule width can be used as a field indicator for high seed weight. Langham (2017a) reported that the coefficient (r) of the correlation of the capsule width and the seed weight per capsule was 0.347. However, the capsule size is not consistent with the high yield. The capsule length increases rapidly in 9–15 days after flowering (DAF). Besides the genotype, the environments of moisture, fertility, and population also affect the development of capsules and seeds (Langham 2018d).

2.2.5.4 Shatter Resistance of Capsule

In most sesame germplasm accessions, capsules are dehiscent with low shatter resistance. The shatter resistance level of a capsule is considered the most important yield component in sesame now, as the indehiscent or non-dehiscent varieties with high shatter resistance can reduce the harvest loss of mature seeds, especially for mechanized production. DG and DR Langham found nine traits affect shattering: capsule opening, capsule split, capsule constriction,

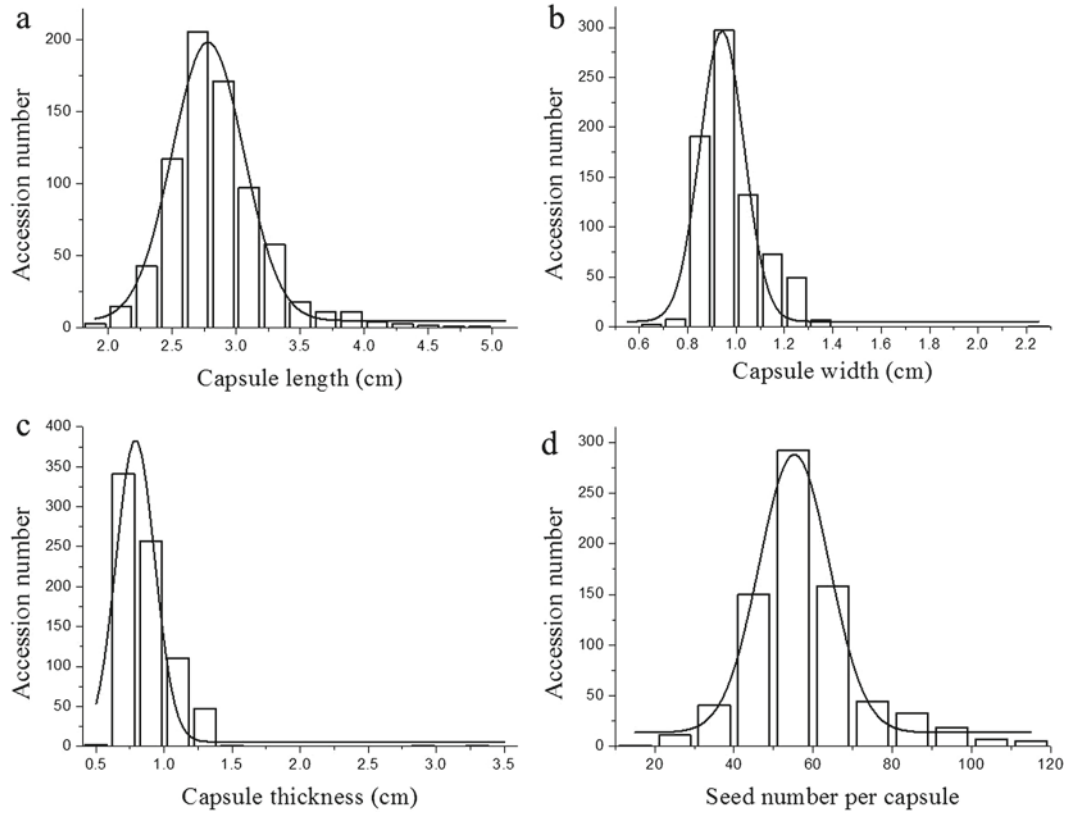


Fig. 2.27 Variation of capsule size in sesame germplasm accessions. **a** Capsule length variation. **b** Capsule width variation. **c** Capsule thickness variation. **d** Seed number per capsule. (Figures provided by Hongmei Miao)

membrane completeness, placenta attachment, capsule membrane attachment, tip roll back, capsule position on stem, and seed blocks (Fig. 2.28a–g).

To reflect the shatter resistance of sesame capsule, Langham (2017a) provides an evaluation method and quantifies the shatter resistance level with the 11 ratings. The measuring procedures are listed as the following: (1) Harvest two capsules from the middle of the main stem from five plants, when the capsules have reached physiological maturity. For triple capsule lines, one central capsule and the one adjacent axillary capsule are chosen. (2) Put the capsules into a container to dry and retain all of the seed. (3) After the capsules are dry, invert the capsules and allow seed to flow out. (4) Collect the seeds out of capsules and the seeds still in capsules, weigh the two groups of seeds, and calculate the percentage of retained seeds. In many shattering cultivars, the capsules will lose seed while still upright, which reflects the loss while the plants are upright in a shock drying.

The shatter resistance standard includes: 0 = Segregating or mixed; 1 = SUS (Super-shattering)—visual seed retention <25%; 2 = SHA (Shattering)—visual seed retention <50%; 3 = SSH (Semi-shattering)—visual seed retention at 50–75%; 4 = SR (Shatter resistant)—visual seed retention >75% without indehiscent or seamless alleles; 5 = ID (Indehiscent)—homozygous indehiscent alleles with closed capsule; 6 = IDO (Indehiscent open)—homozygous indehiscent alleles with capsule open at tip; 7 = GS (Seamless)—homozygous seamless alleles with closed capsule; 8 = GSO (Seamless) (homozygous seamless alleles with capsule open at tip); 9 = C1 (presence of C + 1 alleles with

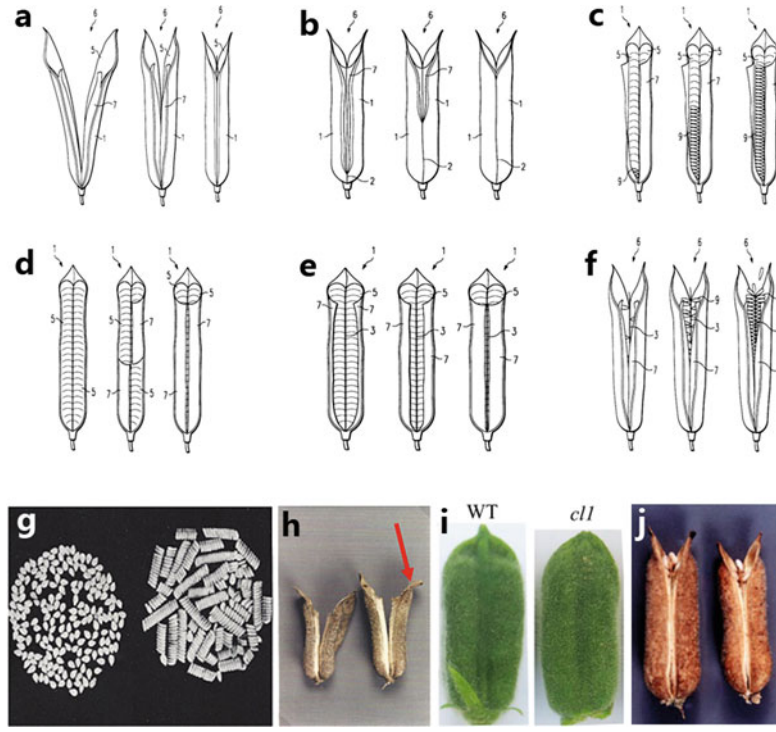


Fig. 2.28 Morphology of sesame capsule. **a** Capsule opening. **b** Capsule split. **c** Capsule constriction. **d** Membrane completeness. **e** Placenta attachment. **f** Capsule membrane attachment. **g** Seed blocks. **h** Tip roll back (indicated by red arrow). **i** Indehiscent capsule of *cl*1 mutant. **j** Non-dehiscent capsule. (Photos a–h and j supplied by Ray Langham; photo i supplied by Hongmei Miao). WT indicates the wild type

closed capsule); 10 = C1O (presence of C + 1 alleles with capsule open at tip); and 99 = Other (Langham 2017a).

A natural mutant (named *cl*1) with indehiscent capsules and curly leaves was found in 1940s (Langham 1946; Langham and Rodriguez 1946) (Fig. 2.28i). The shatter resistance level is 100%. Compared with the natural dehiscent capsule, the indehiscent capsule is entirely closed, and the carpels did not split naturally after maturity or desiccation. The indehiscent trait in *cl*1 is controlled by single gene pair and is not affected by environment (Zhang et al. 2018). The indehiscent line is not suitable for production, as the capsules are hard to break, and if sufficient force is applied to break them open, the seed is damaged leading to free fatty acids and eventually rancidity. Cagirgan (2001) isolated eight indehiscent mutants from irradiating 4 Turkish varieties. Mussi et al. (2016) using 600 Gy found out of 85 plants, two were indehiscent and 13 were semi-dehiscent.

In 1988, Langham released direct harvest varieties that open at the tip, but still retain the majority of the seeds while drought. This was the first varieties that would allow the plants to remain in the field for as much as 50 days after maturity to dry down to 6% seed moisture. In 1997, Langham released nondehiscent varieties that were appreciably more shatter-resistant with less seed loss than the direct harvest varieties (Langham 2000). The non-dehiscent trait was added to hundreds of different genetic backgrounds. Non-dehiscent variety breeding was a giant step ahead in the history of the sesame breeding of the world and improved the mechanization of sesame, although there were still some shatter resistance issues.

In addition, environmental conditions such as rainfall, temperature, and wind can affect the shatter resistance of sesame capsule. In the USA, sesame is harvested in the fall when there is more rainfall. Rainfalls would result in dews and fogs. Under the environment, the hygroscopic capsule would absorb moisture and closes during the night and then opens in the sun in the morning. This constant closing and opening weaken the shatter-resistant mechanisms. If there was

considerable rain, the soil would be too muddy for the combines to enter the fields without leaving ruts and compacting the soil. In general, the breeding nurseries were left standing for as much as 2–4 months after the plants were dry enough to combine to take more notes and make more selections. Under the rains or at times 50 KPH winds, the sesame varieties present the amount of shatter resistance. In 2007, Langham established a new standard for evaluating the shatter resistance of a variety: the plants have to hold 85% of the seeds in a capsule for one month after they are dry, and it has rained at least 30 mm. The varieties that pass this threshold are termed improved non-dehiscent varieties (Langham 2011b, 2013).

Another natural mutant (named seamless) was found in 1986 by DG and DR Langham (Fig. 2.29). The capsule did not open at maturity or desiccation. It had the same problems as the indehiscent capsules—too difficult to open without damaging the seed. Wongyai and Chowchong (2003) in Thailand discovered the C + 1 non-shattering allele. C + 1 is a stabilized variety derived from a cross between a shattering line (KUds6111) and a direct harvest line (Sesaco 20). Upon drydown, the capsules do not open, but release the seeds easily when threshed with a soybean thresher. The hold is different from the S20 hold where the capsules open at the tip when they dry down. The capsules hold the seeds through placenta attachment. When the C + 1 capsules open, there is no placenta attachment, and the seeds easily flow out of the capsule. In the USA where the plants are left standing in the field until they dry down enough to result in 6% moisture seed in the combine, the C + 1 capsules will eventually open before the population is dry enough for harvest and release the seeds (Fig. 2.29). In a wind, with no placenta attachment, most of the seed comes out of the capsules and would be lost. On the other hand, if the plants are cut at physiological maturity and shocked as in most of the sesame production regions in the world, the capsules do not open even when the plants are fully dry until they are threshed.

Fig. 2.29 Two other types of non-shattering alleles. Left photo: seamless capsule where there are two carpels, but they do not form a false membrane to allow the capsule to split at drydown. Right photo: C + 1 capsule where the capsule does not open unless left in the field long after drydown. Photographs provided by Ray Langham

2.2.6 Sesame Seed

2.2.6.1 Seed Morphology and Size

As the sesame capsule grows, the seeds mature in 4–6 weeks. Sesame seeds are attached to the placenta and arranged in rows within the locules of capsule (Ashri 2007). Sesame seeds present various shapes such as orbicular, eclipse, elongated, and column-shaped (Fig. 2.30). Hiltebrandt (1932) distinguished two types in respect of shape of seeds-elongated and orbicular. Joshi (1961) and Ram (1930) showed the seed shapes and colors and found the surface of both white and black sesame seeds was possibly smooth or rough. For the most cultivated sesame varieties, the seed surface is smooth. Some white coat seeds are with web waves or spots on the surface and a line appears in the middle of the surface (Fig. 2.30a). The tip of the sesame seed always presents a black spot, which is the dried placenta attachment (Fig. 2.30). The seeds of the wild species have the rough surface. Only the wild species *S. alatum* forms winged seeds (Fig. 2.30d).

Sesame seed belongs to small-sized type. Langham (2018d) provides the evaluation of seed size with 6 scales using hundred seed weight: 0 = segregating; 3 = small (<0.25 g); 4 = (0.25–0.29 g); 5 = medium (0.30–0.34 g);

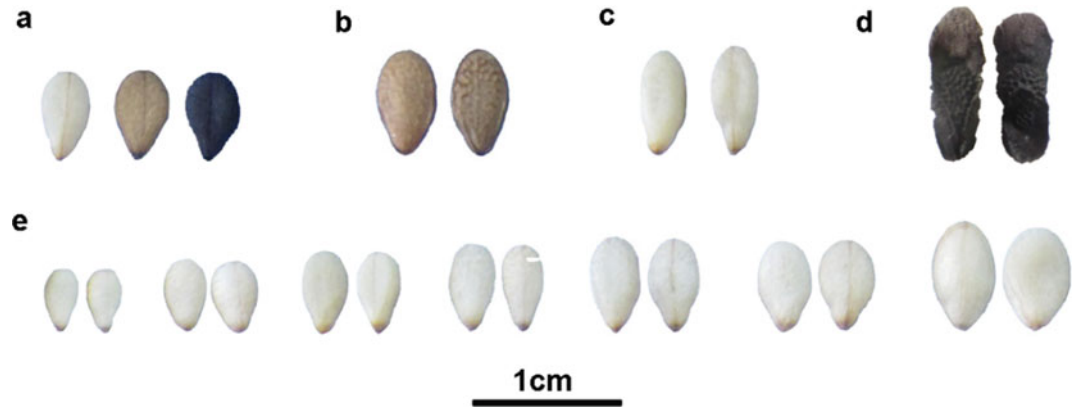


Fig. 2.30 Morphology of sesame seed. **a** Cultivated sesame seeds with stripe in the middle of the surface. **b** Oval shaped seed. **c** Elongated shaped seeds. **d** Winged seeds of *S. alatum*. **e** Seed size variation of sesame varieties. (Photographs provided by Hongmei Miao)

6 = (0.35–0.39 g); and 7 = large seeds (>0.39 g). According to comparison results of 244 samples, the correlation (*r*) between 100-seed weight and the seed size was 0.469 (Langham 2003).

Of the 763 worldwide sesame germplasm accessions cultured in China in 2017, 1,000 seed weight ranged from 0.79 g to 4.47 g (Fig. 2.31a) (Unpublished data, Haiyang Zhang). In China, seeds with the thousand seed weight of more than 3.5 g are regarded as large seeds. Of the 763 accessions, 38 (5.0%) were large-seed varieties. Langham (2018d) reported the international current capsule length range is 0.7–5.2 g. Meanwhile, in the 763 accessions, the length and width of seed varied from 1.92 to 3.58 mm, and 1.14 to 1.97 mm, respectively (Fig. 2.31b, c). The seed area varies from 2.12–5.37 mm^2 (Fig. 2.31d). Correlation analysis indicates that the 1,000 seed weight has the most significant correlation ($P < 0.0001$) with seed area ($r = 0.668$***), seed length ($r = 0.487$***) and seed width ($r = 0.566$***), respectively (data not shown).

In other publications, other measurements of the seeds have been made as follow (Langham 2018d).

Seed length: In mm accurate to 0.01. Current known range: 1.92–4.60 mm.
Seed width: In mm accurate to 0.01. Current known range: 1.00–3.31 mm.
Seed thickness: In mm accurate to 0.01. Current known range: 0.11–1.69 mm
Seed geometric mean diameter: In mm accurate to 0.01. Current known range: 0.81–3.91 mm.
Seed arithmetic diameter: In mm accurate to 0.01 mm. Current known range: 1.29–1.91 mm.
Seed sphericity: In mm^2 accurate to 0.01. Current known range: 0.575–0.64 mm^2. In percentage accurate to 0.01. Current known range: 62.84%.
Seed bulk density: In kg/m^3 accurate to .001. Current known range: 0.640 kg/m^3.
Seed surface area: In mm^2 accurate to 0.01. Current known range: 5.84–10.2 mm^2.
Seed flat surface area: In mm^2 accurate to 0.01. Current known range: 3.24 mm^2.
Seed transverse surface area: In mm^2 accurate to 0.01. Current known range: 1.22 mm^2.
Seed aspect ratio: In percentage accurate to 0.01. Current known range: 61.50–75.00%.

2.2.6.2 Sesame Seed Coat Color

Seed coat color is an important agronomic trait in sesame. The highest polymorphism can be observed for seed coat color ranging from white to black through all intermediate colors (Prasad and Gangopadhyay 2011; Pandey et al. 2013). The natural color of mature sesame seeds varies from black, intermediate colors (e.g., gray, brown, golden, yellow, and light white) to white (Zhang et al. 2013).In the description of the Anon (2004), the seed color of sesame is divided

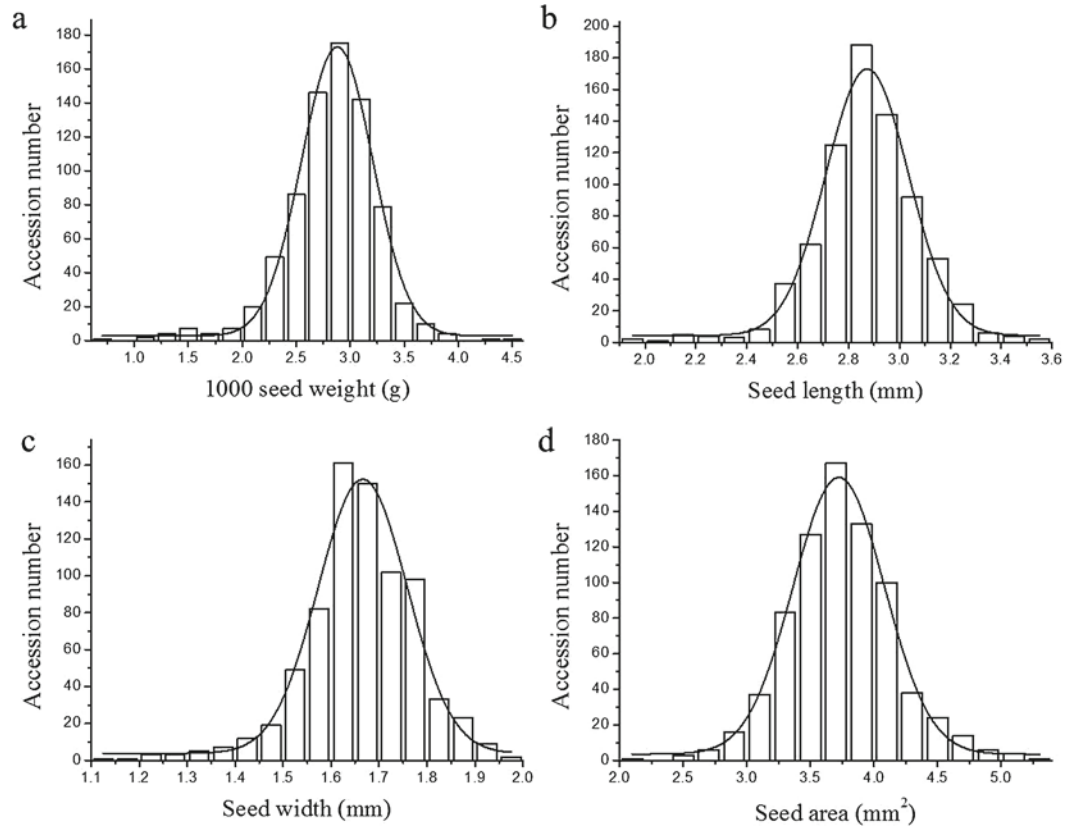


Fig. 2.31 Variation of seed shape and size in 763 sesame germplasm accessions. **a** 1000 seed weight variation. **b** Seed length variation. **c** Seed width variation. **d** Seed area variation (Provided by Hongmei Miao)

into white, cream, beige, light brown, medium brown, dark brown, brick red, tan, olive, grey, dull black, and bright black. White seeded sesame is more popular in Asia and mainly used to produce fried sesame seeds and tahini. Meanwhile, Asian people like to eat black sesame. Especially in China, black sesame is applied as a kind of Chinese medicine. Compared with the white seeds, black sesame seeds usually have higher ash and carbohydrate content, but lower protein, oil, and moisture ratios (Kanu 2011). In addition, there is a commercial sesame variety known as 'Turkey gold' from Turkey. Cagirgan I explained that the seed color of the variety ranges from dark buff to brown and is affected by environment (personal communication, 2013). In sesame, the seed coat color is associated with seed quality traits, disease resistance, and the species evolution trace (Shahidi et al. 2006; El-Bramawy et al. 2008; Zhang et al. 2013). Black varieties generally produce lower yields, probably because of the less effort to improve black varieties.

Based on the variation of RGB (red, green, and blue color) values of seed picture collected using the color capture tool, Zhang et al. (2013) divided the seed coat color into 14 groups (Fig. 2.32). The RGB values range from 20 to 150.

Meanwhile, the color value of sesame seeds can be calculated by L*, a*, and b* values, which are collected using seed color monitor (colorflex EZ, USA). L* represents the color brightness, a* represents the color red and green, and b* represents yellow and blue (Imtiaz et al.

Fig. 2.32 Seed coat color variation in sesame populations. 14 pieces (**a**–**n**) of seed color pictures represented the corresponding RGB values within 14 grades of 20–150. The black seeded parent presents black (**a**) and the white seeded parent is white (**n**). **a** RGB value of 20 ~ 30. **b** 30 ~ 40. **c** 40 ~ 50. **d** 50 ~ 60. **e** 60 ~ 70. **f** 70 ~ 80. **g** 80 ~ 90. **h** 90 ~ 100. **i** 100 ~ 110. **j** 110 ~ 120. **k** 120 ~ 130. **l** 130 ~ 140. **m** 140 ~ 150. **n** 150 ~ . (Cited from Zhang et al. (2013))

Table 2.1 Evaluation standard of seed coat color variation in sesame[a]

Scale	ΔE*ab value	Color variation type
1	$0.0 < \Delta E^*ab \leq 0.5$	Faint difference
2	$0.5 < \Delta E^*ab \leq 3.0$	Difference
3	$3.0 < \Delta E^*ab \leq 6.0$	Significant difference
4	$\Delta E^*ab > 12.0$	Different color

[a]Evaluation standard of seed coat color variation in sesame is referenced the standard of Li (2001) (cited from Yuan et al. (2018))

2008). The equation of $\Delta E = \sqrt{\Delta a^{*2} + \Delta b^{*2} + \Delta L^{*2}}$ is used to evaluate the color variation between two samples or treatments (Yuan et al. 2018; Table 2.1).

Under the waterlogging treatment of 60 h, theΔE*ab value of six sesame varieties (4 white seeded and 2 black seeded) varies from 2.16 to 7.28. The ΔE*ab value variation reflects that the seed color of a variety significantly changes under waterlogging, which is in line with the observation of the seed coat color variation using naked eyes (Yuan et al. 2018).

2.2.6.3 Seed Flavor

Seed flavor is an extremely subjective trait in sesame. The seed flavor comprises of sweetness, astringency, bitterness, sourness, dusty-taste, first-taste, after-taste, accent of taste, and total taste for raw and roasting samples (Takada and Uno 2001). Langham (2018d) lists the values of 4 ratings: 0 = Segregating or mixed; 3 = Bitter; 5 = Neutral; and 7 = Sweet. Ashri (2007) reported there is no information on the genetic control of the taste components for the complicated evaluation and the genetic background. Thus, sesame breeding for sweetness is difficult and complicated. Moreover, the seed flavor can be affected by environment. The amount of rainfall, dews, and humidity can result in the seed taste a bit musky (Langham 2017b). Takada and Uno (2001) rated seed on a −2 to +2 scale on the taste. Bennett MR (personal communication, 2016) stated that the taste of seed is associated with the level of phenols and oxalic acid normally found in the seed testa. Dark colored seeds have phenols and higher oxalic acid content than white seeds. Bennett and Lang (1991) used the following rating of flavor: 0 = tasteless and 1–15 scale with 15 bitter. Shahidi et al. (2010) identified over 240 volatiles in sesame that contribute to flavor to include pyrazines, furans, nitrogen-containing compounds, aldehydes and ketones, alcohols, acids, esters, and lactones, hydrocarbons, phenols, and sulfur-containing compounds. The flavor depends on methods of processing to include temperatures and duration. There is still

disagreement on the most important compounds, but there is agreement that the pyrazines (particularly alkylpyrazines) are responsible for the desirable flavor of sesame.

2.3 Botanical Traits of Wild Species

Since the nineteenth century, many botanists reported the various numbers of the wild species in *Sesamum* (Joshi 1961; Bedigian 2010b; Nimmakayala et al. 2011). Joshi (1961) listed 36 species of *Sesamum* according to the index Kewensis. Some wild species has high economic and academic position in *Sesamum*. For example, *S. alatum* and *S. angustifolium* are used as cooked vegetables especially in eastern Africa. Meanwhile, *S. angustifolium* has high adaptations for drought resistance and can be used for genetics and breeding research. *S. radiatum* is immune to *Fusarium oxsporum* f. sp. *sesami*, and has been used for disease resistance and plant-pathogen interaction studies in sesame. We here summarize the botanic characters of the six wild species (*S. alatum* Thonn, *S. latifollum* Gillet, *S. angolense* Welw, *S. calycinum* Welw, *S. angustifolium* Oliver Engl., *S. radiatum* Schumach. & Thonn.), as these species are being applied for genome sequencing and genomics research in sesame. The basic agronomic traits of the wild species are list in Table 2.2.

2.3.1 *S. Alatum* Thonn.

S. alatum ($2n = 26$) distributes in Asia and Africa. *S. alatum* is an erect annual herb up to 1.5 m tall, with simple or sparsely branched stem. Low leaves are divided almost to the base, and the up leaves linear (Fig. 2.33). Corolla tube is pink or bright red. Dark spot presents at the throat of corolla. Capsule is slim and long with beak. The capsule is bicarpella with small size of seeds. Seed presents a large and long wing at apex. Testa color varies from pale to dark brown (Fig. 2.33e). Compared to the plants cultivated in India, the plant height of *S. alatum* cultivated in China is lower to 1.4 m.

2.3.2 *Sesamum Angolense* Welw

S. angolense Welw. ($2n = 26$) distributes in Africa (https://plants.jstor.org). *S. angolense* is a shrubby perennial herb high to 3 m (Fig. 2.34). Leaves are subsellile or shortly petiolate. Flowers are pink, red, purple, or pale mauve with deeper markings. Corolla is large and the length ranges from 3.5 cm to 7 cm. Single flower per leaf axil (Fig. 2.34b). Yellow nectary can be seen on the leaf axil (Fig. 2.34c). No winged seeds are broad with rough texture (Fig. 2.34d).

2.3.3 *Sesamum Latifolium* Gillet.

S. latifolium Gillet. ($2n = 32$) is erect high to more than 2 m (Fig. 2.35a) (https://plants.jstor.org). The stem has pubescent with brownish hairs. Leaves are heteromorphic. The lower leaves are large, ovate-cordate, or 3-lobed with long petiolate (Fig. 2.35b). Upper leaves are smaller and ovate to ovate-lanceolate. The flowers are pale pink or pinkish mauve (Fig. 2.35c). There are purple stripes on the corolla lop. No winged seeds are broad with sharp-angled margins (Fig. 2.35e).

2.3.4 *Sesamum Calycinum* Welw

S. calycinum Welw. ($2n = 32$) is a semi-erect or prostrate herb (Fig. 2.36). Branches are slender and upwards. Leaves are entire, linear, and subsessile. Corolla is pale rose. Capsules are long with long beak. Capsule length is about 2.96 cm, and the width is about 0.4 cm (Table 2.2). Seeds are small and rugose. Thousand seed weight is

Table 2.2 Characterization of the six *Sesamum* species

Trait	Species						
	S. alatum		*S. angolense*	*S. angustifolium*	*S. calycinum*	*S. latifolium*	*S. radiatum*
	Sample (Akola, India)	3651 (Sanya, China)	K16	G01	Ken 8	Ken 1	G02
Plant height (cm)	167.8	158.0	262.0	121.8	152.2	258.4	191.6
Branch number	5.0	2.0	5.5	17.6	11.6	10.8	16.0
Length of lower leaf (cm)	4.15	13.0	20.5	15.2	11.8	16.7	17.4
Length of upper leaf (cm)	5.45	7.9	9.8	8.7	9.7	11.9	9.7
Width of lower leaf (cm)	0.35	1.0	8.0	4.02	3.44	18.1	13.0
Width of upper leaf (cm)	0.3	0.4	3.3	0.84	0.96	5.1	3.56
Root length (cm)	19.4	*ND*	*ND*	*ND*	*ND*	*ND*	*ND*
Root short ratio	0.12	*ND*	*ND*	*ND*	*ND*	*ND*	*ND*
Corolla length (cm)	3.14	*ND*	7.36	3.68	4.4	4.4	4.28
Corolla width (cm)	1.12	*ND*	5.08	2.2	2.68	2.2	3.0
Petiole length of lower leaves (cm)	2.1	4.5	1.5	*ND*	*ND*	*ND*	*ND*
Height of first capsule (cm)		18.0	129.0	46.0	22.1	89.2	58.4
Capsule zone length (cm)		*ND*	*ND*	64.6	120.0	145.8	132.8
Capsule node number		28.0	*ND*	15.2	29.2	21	27
Stamen number		4	4	4	4	4	4
Pistil length (cm)		1.4	3.0	1.88	2.14	1.7	2.04
Capsule number per plant	52.5	*ND*	*ND*	318.2	339.0	178	298.4
Capsule length (cm)	3.74	3.66	2.88	1.87	2.96	3.4	3.2
Capsule width (cm)	0.52	0.38	0.6	0.4	0.4	0.7	1.1
Seed number per capsule	85.0	85.0	135.8	80.4	104.0	106.2	86.6
1000 seed weight (g)	1.1	1.7	0.939	0.528	0.549	1.523	2.171
Oil content (%)		30	17.88	19.16	14.69	11.34	12.93
Protein content (%)		14.7	31.04	21.22	29.11	28.67	28.24

ND not determined

[a]The first column data (Akola, India) is cited from Babu et al. (2016). Other column data (Sanya, China) is supplied by Hongmei Miao. The wild species are cultured in Sanya, China from 2015–2017. The 5 plants per species are chosen for trait investigation. Oil and protein contents of the seeds are measured using chemistry methods (GB/T5512-2008 and GB/T4489.2-2008, in China)

Fig. 2.33 Botanical characters of *S. alatum*. **a** Plantlet image. **b** Basal leaf. **c** Flower. **d** Capsule. **e** Seeds with wings. (Provided by Haiyang Zhang)

less than 0.6 g. Hybridization results indicates that *S. calycinum* can cross fertile to *S. angustifolium* ($2n = 32$).

2.3.5 *Sesamum Angustifolium* Oliver Engl.

S. angustifolium Engl. ($2n = 32$) is erect annual or perennial herb (Fig. 2.37). Stem is simple or branched. Leaves are entire and linear to cultrate. Flowers are pink, purple, or mauve. Spots present within the corolla. Nectary is sessile. Capsule is long with long beak. The capsule size is long with the length of 1.87 cm and the width of 0.4 cm. The thousand seed weight is less than 0.6 g. Both sides of the seeds show transversal ribs (Fig. 2.37e). *S. angustifolium* can cross fertile to *S. calycinum* ($2n = 32$).

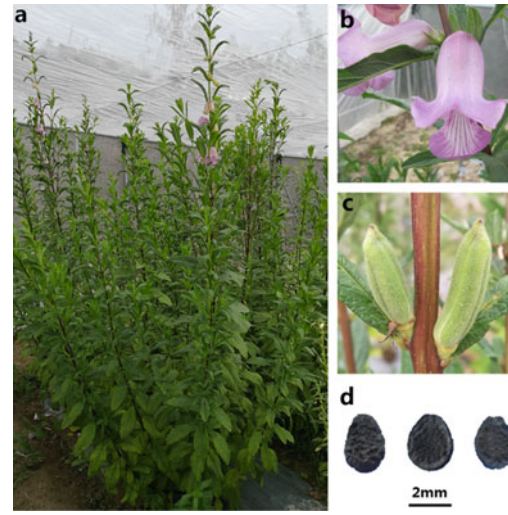


Fig. 2.34 Botanical characters of *S. angolense*. **a** Plantlet image. **b** Flower. **c** Capsule. **d** Seed with wings. (Provided by Haiyang Zhang)

2.3.6 *Sesamum radiatum* Schumach. & Thonn.

S. radiatum Schum. & Thonn. ($2n = 64$) is a perennial or annual herb and distributes in Africa (Dansi et al. 2012). In India, *S. radiatum* is regarded as an invasive species mainly for human activities (Narasimhan 2010). *S. radiatum* is erect up to 120–150 cm (Fig. 2.38). Stem is single or branched. Leaves are simple and stipules are absent. Single flower forms from one leaf axil. Corolla color varies from white, pink to purplish. Capsule is bicarpel with short beak at the apex. Seeds are brown or black with radial sculptures (Dansi et al. 2012) (Fig. 2.38e). To date, *S. radiatum* is cultivated in Africa mainly for nutritional and medicinal properties. The resistance to biotic and abiotic stresses is very high.

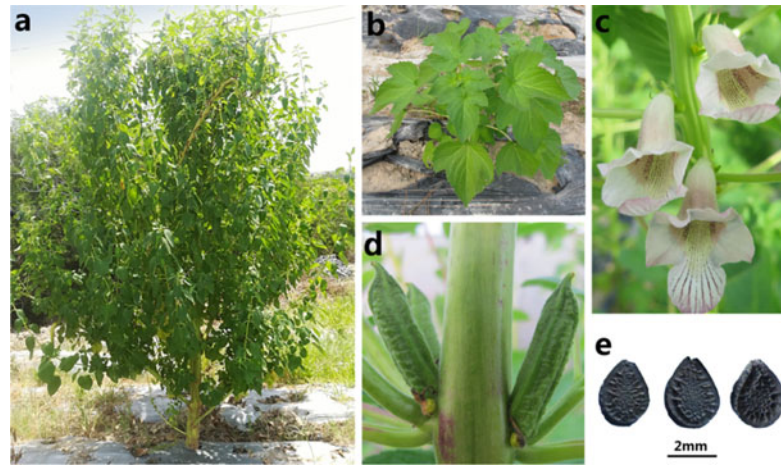


Fig. 2.35 Botanical characters of *S. latifolium*. **a** Plantlet image. **b** Leaf. **c** Flower. **d** Capsule. **e** Seeds with rough surface. (Provided by Haiyang Zhang)

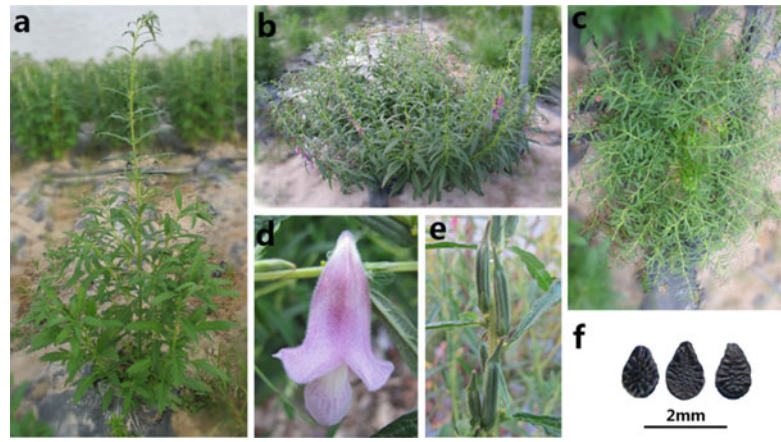


Fig. 2.36 Botanical characters of *S. angustifolium*. **a** Plantlet image. **b** Semi-prostrate plant with many branches. **c** Prostrate plantlet with many branches. **d** Flower. **e** Capsule with beak. **f** Seed with rough surface. (Provided by Haiyang Zhang)

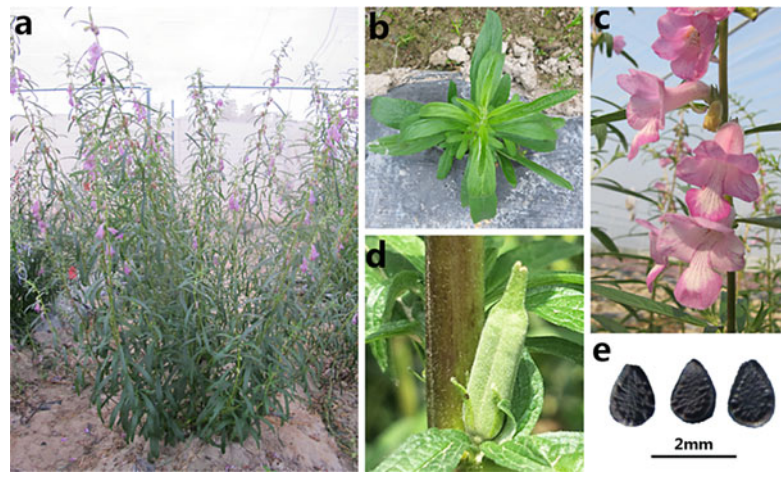


Fig. 2.37 Botanical characters of *S. angustifolium*. **a** Plantlet image. **b** Leaf. **c** Flower. **d** Capsule with beak. **e** Seed with rough surface. (Provided by Haiyang Zhang)

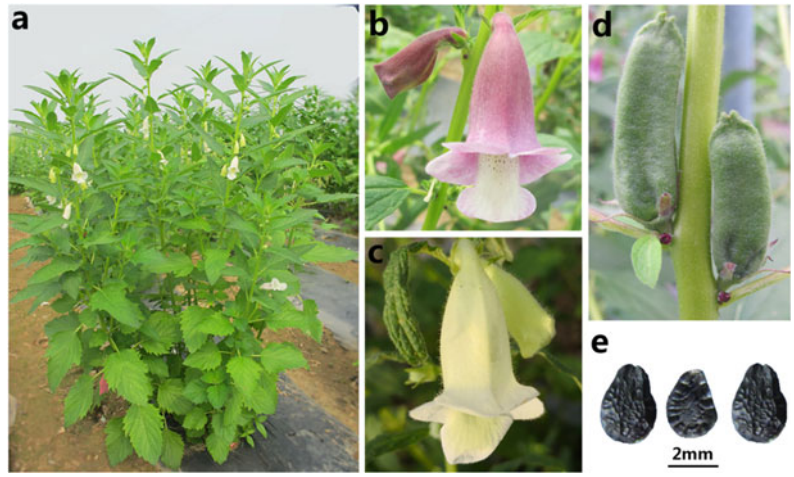


Fig. 2.38 Botanical characters of *S. radiatum*. **a** Plantlet image. **b** Flower with purple corolla. **c** Flower with white corolla. **d** Capsule. **e** Seed with rough surface. (Provided by Haiyang Zhang)

The morphological characters of the wild species are diverse. Therefore, to avoid the ambiguities of the species classification and nomenclature, further cytogenetic and genomic studies are required. Cytogenetic and genomic analyses of the wild species are introduced in the following chapters.

Acknowledgements We would like to thank the famous sesame scientists including Dr. Derald G. Langham from the United States, Prof. Amram Ashri from Israel, and other worldwide sesame scientists for their contribution and donation of the precious germplasm and mutants.

References

Abe A (1919) Goma no nisanno Keisity no Iden Kenkyu Yoho. Agri Rept Formosa 153:15–18

Anon (1998) Sesame certification standards. Arizona Crop Improvement Association Handbook, p 3

Anon (2004) Descriptors for sesame. International Plant Genetic Resource Institute, ISBN 92–9043–632–8, p 76

Ashri A (1998) Sesame breeding. In: Janick J (ed) Plant breeding reviews. Wiley, Israel, pp 179–228

Ashri A (2007) Sesame (*Sesamum indicum* L.). In: Singh RJ (ed) Genetic resources chromosome engineering, and crop improvement, oilseed crops. CRC Press, Taylor and Francis Group, Boca Raton, FL, USA, pp 231–289

Babu JS, Reddy CDR, Reddy NS (2004) Studies on genetic variability in sesame (*Sesamum indicum* L.). Ann Agric Bio-Res 9:7–11

Baydar H, Turgut I (2000) Studies on genetics and breeding of sesame (*Sesamum indicum* L.) I. Inheritance of the characters determining the plant type. Turk J Biol 24:503–512

Bedigian D (1984) *Sesamum indicum* L. Crop origin, diversity, chemistry and ethnobotany. PhD Thesis, University of Illinois, Champaign-Urbana University Microfilms DA8502071 Dissertation Abstracts International, 45:3410–B

Bedigian D (1985) Is še-giš-ì sesame or flax? Bull Sum Agri 2:159–178

Bedigian D (1988) *Sesamum indicum* L. (Pedaliaceae): Ethnobotany in Sudan, crop diversity, lignans, origin, and related taxa. In: Goldblatt P, Lowry PP (eds) Modern systematic studies in African botany. AETFAT Monographs in Systematic Botany, Missouri, pp 315–321

Bedigian D (2000) Sesame. In: Kiple KF, Ornelas-Kiple CK (eds) The cambridge world history of food. Cambridge University Press, New York, pp 411–421

Bedigian D (2010a) Cultivated sesame, and wild relatives in the genus *Sesamum*. CRC Press, Boca Raton, FL, pp 33–77

Bedigian D (2010b) Chracterization of sesame (*Sesamum indicum* L.) germplasm: a critique. Genet Resour Crop Evol 57:641–647

Bedigian D, Harlan JR (1986) Evidence for cultivation of sesame in the ancient world. Econ Bot 40(2):137–154

Bedigian D, Seigler DS, Harlan JR (1985) Sesamin, sesamolin and the origin of sesame. Biochem Syst Ecol 13(2):133–139

Beech DF, Imrie BC (2001) Breeding for mechanized sesame production in Australia. In: Van Zanten L (ed) Sesame improvements by induced mutations. Proceeding final FAO/IAEA co-ord. research Mtg, IAEA, Vienna, TECDOC-1195, pp 63–70

Bennett MR, Lang M (1991) Effect of split application of nitrogen on growth, capsule number and seed yield of sesame cv. Yori 77. In: Sesame research report, 1989–1990 Wet Season, Katherine. Department of Primary Industry and Fisheries, Technical Report No. 167, pp 21–25

Bhat KV, Babrekar PP, Lakhanpaul S (1999) Study of genetic diversity in Indian and exotic sesame (*Sesamum indicum* L.) germplasm using random amplified polymorphic DNA (RAPD) markers. Euphytica 110:21–33

Bisht IS, Mahajan RK, Loknathan TR, Agrawar RC (1998) Diversity in Indian sesame collection and stratification of germplasm accessions in different diversity groups. Genet Res Crop Evol 45:325–335

Blal AEH, Kamel SM, Mahfouz HM, El-Wahed MSA (2012) Impact of pollination and fertilization on sesame production in the reclaimed lands. Ismailia Governorate, Egypt J of Agric Sci 57(3):121–133

Blal AEH, Kamel SM, Mahfouz HM, Said M (2013) Impact of opened, non opened pollination and nitrogen fertilizer on sesame production in the reclaimed lands, Ismalia Governorate. Egypt. Cercetari Agronomice Moldova 56(3):57–68

Brar GS, Ahuja KL (1979) Sesame: its culture, genetics, breeding and biochemistry. AnnRev Plant Sci 1:245–313

Brummitt RK (2005) Report of the Committee for Spermatophyta: 48. Taxon 48(2):359–371

Cagirgan MI (2001) Mutation techniques in sesame (*Sesamum indicum* L.) for intensive management: confirmed mutant. In: Van Zanten (ed) Sesame Improvement by Induced Mutations, L. IAEA-TECDOC-1195, IAEA, Vienna, pp 31–40

Culp T (1960a) Inheritance of papershell capsules, capsule number, and plant color in sesame. J Hered 51(3):101–103

Culp TW (1960b) Inheritance of plant height and capsule length in sesame. Agron J 52(2):101–103

Dansi A, Etèka CA, Adéoti K, Orkwor GC, Ahohuendo Y et al (2012) Black benniseed (*Sesamum radiatum* Schum. et Thonn.) cultivated as leafy vegetable in Benin. Genet Resour Crop Evol 59:955–964

El-Bramawy MAES, El-Hendawy SES, Amin Shaban WI (2008) Assessing the suitability of morphological and phenological traits to screen sesame genotypes for *Fusarium* wilt and charcoal rot disease resistance. J Plant Protect Res 48(4):397–410

Falusi OA (2007) Studies on cross and self-compatibility in sesame. J Arid Agri 17:11–13

Frary A, Tekin P, Celik I, Furat S, Uzun B et al (2015) Morphological and molecular diversity in Turkish sesame germplasm and core set selection. Crop Sci 55:1–10

Fuller D (2000) Fifty years of archaeobotanical studies in India: laying a solid foundation. In: Settar S, Korisettar R (eds) Indian archaeology in retrospect III. Archaeology and Interac-tive Disciplines. Manohar, New Delhi, India, pp 247–363

Fuller DQ (2003) Further evidence on the Prehistory of sesame. Asian Agri-History 7(2):127–137

Fuller D, Madella M (2000) Issues in Harappan Archaeobotany. Retrospect and Prospect. In: Settar S and Korisettar R (eds), Indian archaeology in retrospect II. Protohistory. Manohar, New Delhi, India, pp 317–390

Ghosh BN (1955) Photoperiodic response in til (*Sesamum indicum* Linn.). Curr Sci 24:170

Guo W, Miao HM, Zhang TD, Wei LB, Li C et al (2014) Cytological observation of partial sterility of nucleic male sterile line ms86-1 in sesame (*Sesamum indicum*). Chinese Bulletin of Botany 49(1):87–97

Hanano S, Goto K (2011) Arabidopsis TERMINAL FLOWER1 is involved in the regulation of flowering time and inflorescence development through transcriptional repression. Plant Cell 23:3172–3184

Hiltebrandt VM (1932) Sesame (*Sesamum indicum* L). Bull Appl Bot Genet and Pl Breed 9:109–114

Hiremath SC, Patil CG (1999) Genome homology and the putative progenitor of sesame. J Cytol Genet 34:69–74

Howard A, Howard GLC, Khan AR (1919) Studies in the pollination of Indian crops. J Mem Dept Agric Ind Bot Ser 10:195–220

Imtiaz M, Ogbonnaya FC, Oman J, Ginkel MV (2008) Characterization of quantitative trait loci controlling genetic variation for preharvest sprouting in synthetic backcross-derived wheat lines. Genetics 178:1725–1736

John CM (1934) Inheritance studies in gingelly, *Sesamum indicum*. Assoc Econ Biol Proc 2:33–40

Joshi AB (1961) Sesamum. Indian Central Oilseed Committee Hyderabad, India, p 109

Kamel S, Blal A, Mahfouz H, Said M (2013) The most common insect pollinator species on sesame crop (*Sesamum indicum* L.) in Ismailia Governorate, Egypt. Arthropods 2(2):66–74

Kanu PJ (2011) Biochemical analysis of black and white sesame seeds from China. Amer J Biochem Mol Biol 1:145–157

Khidir MO (1972) Natural cross-fertilisation in sesame under Sudan conditions. Exp Agri 8:55–59

Khidir MO (1973) Genetic studies in sesame. II. Inheritance of flower colour and number of locules per pod. Wxp Agri 9(4):361–364

Kobayashi T (1986) Some genetic patterns for the main characteristics of the cultivated sesame. In: Fernandez MJ (ed) Sesame and safflower newsletter. CIDA, Cordova, Spain, pp 23–32

Kobayashi T (1991) Cytogenetics of Sesame (*Sesamum indicum*). Dev Plant Genet Breed 2:581–592

Kumar LS, Gokhale AV (1952) Natural cross-pollination in *Sesamum orientale* L. in the Bombay Deccan. Poona Agric Coll Mag 43:99–101

Langham DG (1944) Natural and controlled pollination in sesame. J Hered 35:255–256

Langham DG (1945a) Genetics of sesame. J Hered 36 (5):135–142

Langham DG (1945b) Genetics of sesame II. Inheritance of seed pod number, aphid resistance, "yellow-leaf" and wrinkled leaves. J Hered 36(8):245–253

Langham DG (1946) Genetics of sesame III. Open sesame and mottled leaf. J Hered 37:149–152

Langham DG (1947a) Genetics of sesame IV. Some genetic variations in the color of the sesame flower. J Hered 38:221–224

Langham DG (1947b) Genetics of sesame V. Some morphological differences of the sesame flower (*Sesamum indicum* L.). J Hered 38:347–352
Langham DG (1947c) Herencia del tipo de capsula en el ajonjoli (*Sesamum indicum* L.). Publicacion del Ministerio de Agriculturay CriaCir. 24, p 7
Langham DR (2000) Method for making non-dehiscent sesame. United States Patent 6,100,452, p 22
Langham DR (2001) Shatter resistance in sesame. In: Van Zanten L (ed) Sesame improvements by Induced mutations, proceeding final FAO/IAEA co-ord research Mtng, IAEA-TECDOC-1195, pp 51–61
Langham DR (2003) Seed size in sesame. Paper shared with sesame researchers from Australia, Israel, S. Korea, Thailand, and USA, ResearchGate.com
Langham DR (2007) Phenology of sesame. In: Janick and Whipkey A (eds) Issues in new crops and new uses. ASHS Press, Alexandria, VA, pp 144–182
Langham DR (2008) Growth and development of sesame, American Sesame Grower Association, San Antonio, TX, p 44
Langham DR (2011a) Non-dehiscent pygmy sesame variety Sesaco 70. United States Patent 8,058,503 B1, p 42
Langham DR (2011b) Non-dehiscent sesame. United States patent 8,080,707, p 21
Langham DR (2013) Method for breeding improved non-dehiscent sesame. United States patent 8,581,028 B2, p 20
Langham DR (2014) Pygmy sesame plants for mechanical harvest. United States patent 8,664,472, p 30
Langham DR (2017a) VII Capsule descriptors of sesame (*Sesamum indicum* L.). ResearchGate, p 263
Langham DR (2017b) VIII Capsule zone descriptors of sesame (*Sesamum indicum* L.). ResearchGate, p 276
Langham DR (2018a) IV Sesame root and stem descriptors (*Sesamum indicum* L.) Working paper 1. ResearchGate, p 51
Langham DR (2018b) V Sesame leaf descriptors (*Sesamum indicum* L.), Working paper 1. ResearchGate, p 77
Langham DR (2018c) VI Sesame flower descriptors (*Sesamum indicum* L.), Working paper 1. ResearchGate, p 181
Langham DR (2018d) X Sesame seed descriptors (*Sesamum indicum* L.), Working paper 1. ResearchGate, p 141
Langham DR (2019) Sesame plant descriptors (*Sesamum indicum* L.), Working paper 1. (In manuscript)
Langham DG, Rodriguez M (1946) Abrete sesamo: Ajonjoli (*Sesamum indicum*) que no pierde semillas. Circ. 17, Publ. Ministerio de Agricultura y Cria, Maracay, Venezuela.
Langham DR, Wiemers T (2002) Progress in mechanizing sesame in the US through breeding. In: Janick J (ed) Trends in new crops and new uses. ASHS Press, Alexandria, VA, pp 157–173
Li K (1987) The anatomical structure of the vegetative organ of the *Sesamum indicum*: II. The leaf anatomy structure. J Henan Norm Univ 53 (1):89–95
Monteverde GE, Quijada P (1994) Herencia del color de la semilla y numero de lóculos por cápsula en tres cultivares de ajonjolí *Sesamum indicum* L. Agron Trop 44(3):513–527
Li S (2001) Physical Properties of Foods. China Agriculture Press Beijing, China, p 335
Mahajan RK, Bisht I, Dhillon BS (2007) Establishment of a core collection of word sesame (*Sesamum indicum* L.) germplasm accessions. SABRAO J Breed Genet 39(1):53–64
Mahfouz HM, Kamel SM, Belal AH, Said M (2012) Pollinators visiting sesame (*Sesamum indicum* L.) seed crop with reference to foraging activity of some bee species. Cercetaqri Agronomica in Moldova 45 (2):49–55
Mei H, Liu Y, Du Z, Wu K, Cui CQ et al (2017) High-density genetic map construction and gene mapping of basal branching habit and number of flowers per leaf axil in sesame. Front Plant Sci 8:636
Melzer S, Lens F, Gennen J, Vanneste S, Rohde A et al (2008) Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana*. Nat Genet 40:1489–1492
Mohammad A, Alam Z (1933) Types of *Sesamum indicum* D.C. in the Punjab. Indian J Agric Sci 3:897–911
Montilla D, Cedeno T (1991) Efecto de la polinizacion por abejas (*Apis mellifera* L.) sobre el rendimiento de cultivares indehiscentes de ajonjoli (*Sesamum indicum* L.). BIAGRO 3(4):139–144
Muhammed SV, Srinvasalu N, Shandrasekharon NR, Sundaram N, Sivasubramaniam P (1965) Studies on natural cross pollination in *Sesamum orientale* L. Madras Agric J 52:452
Mussi C, Nakayama H, Oviedo de Cristaldo R (2016) Variabilidad fenotipica en poblaciones M_1 de sesamo (*Sesamum indicum* L.) irrdiado con rayos gamma. Cultivos Tropicales 37(especial):74–80
Nanthakumar G, Singh KN, Vaidyanathan P (2000) Relationships between cultivated sesame (*sesamum* sp.) and the wild relatives based on morphological characters, isozymes and RAPD markers. J Genet Breed 54:5–12
Narasimhan D, Arisdason W, Irwin SJ, Gnanasekaran G (2010) Invasive alien plant species of Tamil Nadu. National Seminar on Invasive Alien Species, 29–38
Ngongolo K, Mtoka S, Rubanza CD (2015) Floral visitors and pollinators of sesame (*Sesamum indicum L)* from Kichi Forest to the adjacent local communities' farms. Entomol Appl Sci Lett 2(2):32–39
Nimmakayala P, Perumal R, Mulpuri S, Reddy UK (2011) Sesamum. In: Kole C (ed) Wild crop relatives: genomic and breeding resources, vol Oilseeds. Springer, Berlin, Heidelberg, pp 261–273
Nithya C, Viraktamath S, Patil RK, Vastrad AS, Palakshappa MG (2012) Influence of indigenous bee attractants in enhancing pollination and yield of sesame. Karnataka J Agric Sci 25(4):537–539
Nohara S (1933) Genetical studies on *Sesamum indicum* L. Tokyo Imp Univ Coll Agri J 12:227–386

Padmasundari M, Kamala T, Rao YV (2010) Genetics or locule in *Sesamum indicum* L. Asian J Agric Sci 2 (4):164–167

Pal BP (1934) Recent progress in plant breeding at Pusa. Agric Live-Stk India 4:505–515

Panda P, Sontakke BK, Sarangi PK (1988) Preliminary studies on the effect of bee (*Apis cerana indica* Fab.) pollination on yield of sesamum and niger. Indian Bee J 50:63–64

Panda P, Nanda UK, Mohapatra H, Padhi, (1989) Insect pollination in some oilseed crops in Orissa, India. Indian Bee J 5:97–98

Pandey SK, Das A, Dasgupta T (2013) Genetics of seed coat color in sesame (*Sesamum indicum* L.). Afr J Biotechnol 12:6061–6067

Pashte VV, Shylesha AN (2013a) Pollen and nectar foraging activity of honeybees in *Sesamum*. Indian J Entomol 75(2):124–126

Pashte VV, Shylesha AN (2013b) Pollinators diversity and their abundance on *Sesamum*. Indian J Entomol 75(3):260–262

Pashte VV, Shylesha AN, Badge AS (2015) Effect of honeybee pollination, bee attractants and scents on yield of *Sesamum*. Intl J Ecol Environ Sci 41:87–93

Pathirana R (1994) Natural cross-pollination in sesame (*Sesamum indicum* L.). Plant Breed 112:167–170

Patil BS, Viraktamath S, Lingappa S, Giraddi RS, Parameshwarappa K et al (2000) Effect of Bee-Q and Bee-here on pollinators and yield of Sesamum. Insect Environ 5:151–152

Powell MA (1991) Epistemology and Sumerian agriculture: the strange case of sesame and linseed. Aula Orientalis 9:155–164

Prasad R, Gangopadhyay G (2011) Phenomic analyses of Indian and exotic accession of sesame (*Sesamum indicum* L.). J Plant Breed Crop Sci 3:336–352

Ram K (1930) Studies in Indian oil seeds. The types of *Sesamum indicum* DC. Indian Dept Agri Mem Bot Ser 18:127–147

Rhind D (1935) A note on photoperiodism in sesamum. Indian J Agric Sci 5:729–736

Richharia RH, Persai DP (1940) Tetraploid til (*Sesamum orientale* L.) from colchicine treatment. Curr Sci 9:542

Said M, Kamel SA, Mahfouz ABH (2013) Impact of insect pollinators on sesame production. Lap Lambert Academic Publishing, Saarbrücken, Germany, p 92

Sajjanar SM, Eswarappa G (2015) Sesame (*Sesamum indicum* L.) crop insect pollinators with special reference to the foraging activity of different species of honeybees. J Agri Vet Sci 8(11):9–14

Saravanan S, Nadarajan N (2003) Combining ability studies in sesame. Sesame & Safflower Newsletter 18:1–6

Shahidi F, Liyana-Pathirana CM, Wall DS (2006) Antioxidant activity of white and black sesame seeds and their hull fractions. Food Chem 99:478–483

Shahidi F, Hughes R, Tan Z (2010) Flavor constituents of sesame. In: Bedigian D (ed) Sesame: the genus *Sesamum*. CRC Press, Taylor & Francis Group, Boca Raton, FL, USA, pp 169–185

Sharma E, Shah TI, Khan F (2014) A review enlightening genetic divergence in *Sesamum indicum* based on morphological and molecular studies. Intl J Agri Crop Sci 7(1):1–9

Sikka SM, Gupta ND (1949) Pollination studies in *Sesamum orientale* L. Indian J Agric Sci 9(1):33–41

Sintim HO, Yeboah-Badu VI (2010) Evaluation of sesame (*Sesamum indicum*) production in Ghana. J Anim Plant Sci 6(3):653–662

Su S, Li R, Lang D, Zhang K, Hao X et al (2016) Microstructure of glandular trichomes on leaf surface of sesame and changes of trichome secretions under drought condition. Acta Agron Sin 42(2):278–294

Sun J, Wei X, Le M, Yan T, Yan X et al (2015) Study on breeding characteristics of *Sesamum indicum* L. II: Ascertain of cross-pollination vectors. Seed 3:36–40

Sun J, Zhang XR, Zhang YX, Wang LH, Huang B (2009) Effects of waterlogging on leaf protective enzyme activities and seed yield of sesame at different growth stages. Chin J Appl Environ Biol 15:790–795

Takada, N, Uno T (2001) Japanese market and Thai black sesame seeds. In: Proceeding of the 2nd national conference sesame, sunflower, castor, and safflower, Aug 16–17, Nakhon Nayok, pp 15–30

Tchuenguem Fohouo FN, Nepide NC (2018) Efficacité pollinisatrice de *Apis mellifera* L. (Hymenoptera: Apidae) sur *Sesamum indicum* (Pedaliaceae) var. graine blanche et lisse à Dang (Ngaoundéré, Cameroun). Intl J Biol Chem Sci 12(1):446–461

vvan Rheenen HA (1980) Aspects of natural cross fertilisation in sesame (*Sesamum indicum* L.). Trop Agri (trinidad) 57:53–59

van Rheenen HA (1981) The desirability of multi-versus four locular capsules in sesame (*Sesamum indicum* L.). In: Ashri A (ed) Sesame: status and improvement, proceedings of expert consultation, FAO Plant Production and Protection Paper Rome, Italy, pp 108–112

van Zanten L (2001) Sesame improvement by induced mutations: Results of the co-ordinated research project and recommendation for future studies. In: Van Zanten L (ed) Sesame improvement by induced mutations 1. Proceeding final FAO/IAEA co-ord. research Mtng, IAEA, Vienna, TECDOC-1195, pp 1–12

Weiss EA (1971) Castor, sesame and safflower. Leonard Hill Books, Weiss London: p 901

Weiss EA (1983) Oilseed crops. Longman Inc., New York, p 660

Weiss EA (2000) Sesame. Oilseed crops. Longman Inc., New York, pp 131–164

Wongyai W, Chowchong S (2003) C plus 1: The new white seed sesame variety with non-shattering. In: Proceedings of the 3rd national conference on sesame, sunflower, castor and safflower, Chiang Mai

Wongyai W, Juttpornpong S (1992) Indirect selection from seed weight in sesame using capsule size as a criteria. Sesame and Safflower Newsletter No 4, FAO, Rome, Italy

Yang XL, Zhang HY, Guo W, Zheng YZ, Miao HM, et al. (2008) Ultrastructure in microspore abortion of

genic male sterile line in Sesame (*Sesamum indicum*, L.). Acta Agron Sin 34(11):1894–1900

Yermanos DM (1980) Sesame. In: Fehr WR, Hadley HH (eds) Hybridization of crop plants. American Society of Agronomy, Inc., and the Crop Science Society of America, Inc, Madison, WI, pp 549–563

Yermanos DM, Edwards RT, Hemstreet SC (1964) Sesame an oilseed crop with potential in California. Calif Agri 18(7):2–4

Yol EB, Uzun, (2012) Geographical patterns of sesame accessions grown under Mediterranean environmental conditions, and establishment of a core collection. Crop Sci 52:2206–2214

Yuan Q, Zhang H, Miao H, Duan Y, Wei Q et al (2018) Effects of water logging stress on the quality of sesame seed and oil product. Acta Agri Bor-Sin 33 (2):202–208

Zhang H, Miao H, Li C, Wei L, Ma Q (2012) Analysis of sesame karyotype and resemblance-near coefficient. Chin Bull Bot 47(6):602–614

Zhang H, Miao H, Wei L, Li C, Zhao R, et al. (2013) Genetic analysis and QTL mapping of seed coat color in sesame (*Sesamum indicum* L.). PLoS One 8(5):e63898

Zhang H, Miao H, Li C, Wei L, Duan Y et al (2016) Ultra-dense SNP genetic map construction and identification of *SiDt* gene controlling the determinate growth habit in *Sesamum indicum* L. Sci Rep 6:31556

Zhang H, Miao H, Wei L, Li C, Duan Y et al (2018) Identification of a SiCL1 gene controlling leaf curling and capsule indehiscence in sesame via cross-population association mapping and genomic variants screening. BMC Plant Biol 18(1):296

Zohary D, Hopf M (1994) Domestication of plants in the old world: the origin of spread of cultivated plants in West Asia, Europe, and the Nile Valley. Oxford University Press, Oxford, p 316

Beneficial Components in Sesame Proteins and Oil

3

Jason T. C. Tzen

Abstract

Sesame seed has been well recognized as a nutritional protein source owing to its richness in sulfur-containing amino acids, particularly methionine. Proteins represent approximately 15–25% dry weight of sesame seed, or 30–50% mass of the defatted sesame cake. Two abundant storage proteins, 11S globulin and 2S albumin, constitute 60–70% and 15–25% of total sesame proteins, respectively. Two gene families separately encoding four 11S globulin and three 2S albumin isoforms were tentatively identified in sesame genome. Amino acid composition of these deduced storage proteins revealed that the richness in methionine and cysteine is attributed mainly to two 2S albumin isoforms and partly to one 11S globulin isoform. Sesame oil, a high-quality edible oil source, comprises mainly triacylglycerols (>95%); the acyl chains esterified to the glycerol backbone are mostly unsaturated essential fatty acids. In comparison with most other edible oils extracted from diverse seeds, sesame oil is extremely stable due to the effective antioxidant activities partly attributed to its abundance of lipid-soluble furofuran lignans, mainly sesamin and sesamolin, which are also demonstrated to possess several biological effects on human health. The oil molecules, including furofuran lignans, in sesame seed are stored in specialized organelles termed oil bodies, which are encapsulated by a layer of proteins on their surface. Three classes of unique proteins, i.e., oleosin, caleosin, and steroleosin, were identified in sesame oil bodies in the past two decades, and gene families encoding three oleosin, two caleosin, and two steroleosin isoforms were found in sesame genome. Particularly, caleosin and steroleosin were named when they were first detected in sesame oil bodies. Therefore, the structural organization of sesame oil body has served as a reference model to study oil bodies in diverse plant species, and the three types of oil body proteins were also subsequently detected in oil bodies of various seeds. Novel techniques for constituting various types of artificial sesame oil bodies with three essential constituents (matrix oil, membrane phospholipid, and oil body protein) have been successfully established, and applications of these artificial oil bodies are actively developed.

J. T. C. Tzen (✉)
Graduate Institute of Biotechnology, National Chung-Hsing University, Taichung, Taiwan
e-mail: tctzen@dragon.nchu.edu.tw

H. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-98098-0_3

3.1 Introduction

Sesame seed, empirically utilized as food additive or key ingredient of various products, has been regarded as a nutritional goldmine. Distinct from most other seeds, sesame is well recognized for two characters beneficial for human health. One is its high quality of nutritive protein composition, i.e., richness in sulfur-containing amino acids. The other is the presence of abundant oil-soluble furofuran lignans, mainly sesamin and sesamolin, which are known to possess several biological activities beneficial for human body. Similar to other oily seeds, sesame seed accumulates abundant storage proteins and oils (mainly triacylglycerols) in protein bodies and oil bodies, respectively (Fig. 3.1). The proteins rich in sulfur-containing amino acids are confined in proteins bodies while furofuran lignans are mixed with triacylglycerols in the lipid matrix of oil bodies.

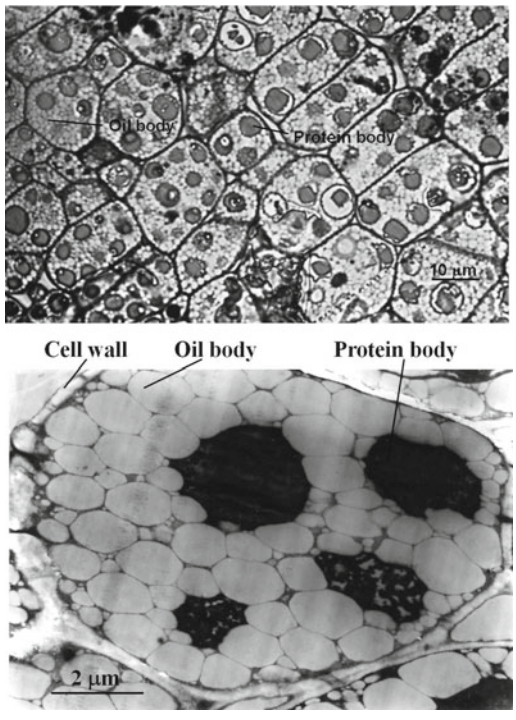


Fig. 3.1 Light microscopy (upper) and electron microscopy (lower) of a mature sesame seed (Peng and Tzen 1998)

3.2 Sesame Storage Proteins

The content of proteins represents approximately 15–25% dry weight of sesame seed, or 30–50% mass of the defatted sesame cake. Amino acid contents of sesame seed vary among varieties and their growing conditions, but not seed coat colors, e.g., white, brown, and black (Namiki 1995). Compared with the standard values recommended for dietary requirement by the Food and Agriculture Organization and the World Health Organization, sesame proteins are slightly lower in lysine but richer in methionine and cysteine. In terms of essential amino acid composition, sesame seed proteins are comparable to those in beef and milk, except for the low content of lysine (Hasegawa et al. 1978). Sesame cake, the remainder after oil extraction, is an adequate supplement of proteins to domestic animals owing to its richness in sulfur-containing amino acids.

3.2.1 Historical Studies on Sesame Proteins

The earliest scientific documentation on the physicochemical properties of sesame proteins appeared in late nineteenth century when protein chemistry was in its infancy and little was known about amino acid constituent. Ritthausen (1880) crystallized sesame seed proteins from defatted sesame cake by saline solution and found a high sulfur content of these proteins by elemental analysis. Approximately half century later, Jones and Gersdorff (1927) modified the extraction procedure of sesame proteins and succeeded in isolating two globulin fractions. The relatively abundant globulin was crystallized and named α-globulin. Thirty years later, Nath and Giri (1957a; b) investigated the physicochemical properties of sesame seed proteins by electrophoresis, and the amino acid composition of sesame seed α-globulin was determined (Nath et al. 1957). By ultracentrifugal analysis, Sinha and Sen (1962) separated sesame seed proteins into four fractions, 2S, 7S, 13S, and 19S

components; and the major fraction, α-globulin, comprising 13S and 19S components, represents 60–70% of total sesame seed proteins. Taking into consideration of homologous proteins from other seeds, sesame α-globulin (13S component) is named 11S globulin thereafter. Molecular cloning of cDNA fragments encoding these well-characterized sesame proteins was not reported until the end of the last century (Tai et al. 1999). The rapid advance of cloning and sequencing techniques has allowed the complete sequencing of two gene families encoding abundant proteins, as well as a comprehensive understanding of the properties of these sesame proteins at molecular level (Hsiao et al. 2006).

3.2.2 The Most Abundant Protein: 11S Globulin

The most abundant protein of sesame seed, the insoluble 11S globulin, constitutes 60–70% of total seed proteins. Purification of sesame 11S globulin was achieved by chromatography using a Sepharose 6B column (Okubo et al. 1979a). Most physicobiochemical properties of the purified sesame 11S globulin were characterized (Table 3.1), including molecular weight, amino acid composition, organization and interaction of subunits, intrinsic viscosity, isoelectric point, specific rotation, absorption coefficient, fluorescence emission maximum, partial specific volume, diffusion coefficient, frictional ratio, quaternary structure, and hydrogen ion equilibrium (Hasegawa et al. 1978; Prakash and Nandi 1978; Lakashmi and Nandi 1979; Nishimura et al. 1979; Okubo et al. 1979b; Plietz et al. 1986; Yuno et al. 1986; Prakash and Narasinga Rao 1990). Gelation properties of sesame 11S globulin play an important role in the preparation of various food products with sesame paste. Compared with soybean 11S globulin, sesame 11S globulin is more hydrophobic; this distinct hydrophobicity results in their differences in solubility, gelation properties, and emulsification capacities. Gelation occurs at pH 7 in the presence of 2% NaCl for soybean protein, but at pH 10 in the presence of 5% NaCl for sesame protein (Hasegawa et al. 1979). To improve gelation properties, succinic and maleic acids were added to sesame protein to increase its hydrophilicity and gelation capability (Hasegawa et al. 1981). Reductive alkylation of ε-amino groups increased the critical concentration for gelation of sesame 11S globulin (Fujino et al. 1981). Moreover, emulsification capacity and stability of sesame 11S globulin were found to be elevated by adding citric acid.

Sesame seed contains several 11S globulin isoforms that assemble randomly to form hexamers of M_r 300–350 kDa (Plietz et al. 1986). Each sesame 11S globulin consists of an acidic subunit (30–40 kDa) and a basic subunit (20–25 kDa) covalently linked by a single disulfide

Table 3.1 Physicochemical properties of 11S globulin and 2S albumin

Property	11S globulin	2S albumin
Sedimentation coefficient ($S_{20,w}$)	12.8S	2S
Diffusion coefficient ($D_{20,w}$)	3.46×10^{-7} cm^2/s	8×10^{-7} cm^2/s
Intrinsic viscosity	0.03 dL/g	4.1 mL/g
Partial specific volume	0.718 mL/g	0.725 mL/g
Absorption maximum	278.5 nm	276 nm
Absorption coefficient	13.0	7.24
Fluorescence emission maximum	328 nm	336 nm
Frictional ratio	1.06	1.3
Molecular weight	361–450 kDa	14–18 kDa

Data from Prakash and Nandi (1978), Nishimura et al. (1979), Rajendran and Prakash (1988)

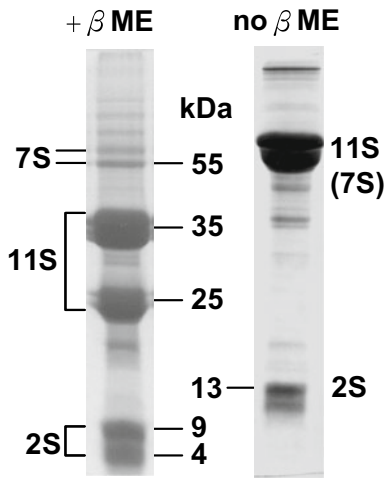


Fig. 3.2 SDS-PAGE of sesame total seed proteins in the presence and absence of β-mercaptoethanol (β-ME)

bond (Tai et al. 1999). These subunit pairs are dissociated after the breakage of their linked disulfide bond in the presence of a reducing agent, such as β-mercaptoethanol (Fig. 3.2). The acidic and basic subunits of 11S globulins are paired specifically as they are processed from a precursor protein of 50–60 kDa, which is cut into two polypeptides posttranslationally after the disulfide bond is first formed between the two subunits.

3.2.3 The Major Soluble Protein: 2S Albumin

The major soluble protein, 2S albumin, which constitutes approximately 15–25% of the total sesame proteins, was first named β-globulin in contrast to the most abundant α-globulin. 2S albumin was purified by precipitation with 30–50% sodium sulfate followed by Sephadex G-100 chromatography with a phosphate buffer containing 0.15 M NaCl, and its physicochemical properties (Table 3.1) and hydrodynamic parameters were determined and calculated (Rajendran and Prakash 1988). Similar to 11S globulin, each 2S albumin isoform consists of a small subunit (4 kDa) and a large subunit (9 kDa) presumably linked by two disulfide bonds (Tai et al. 1999; Fig. 3.2). The small and large subunits of 2S albumin are paired specifically as they are processed from a precursor protein, which is possibly cut into two polypeptides by the same protease responsible for splitting the acidic and basic subunits of 11S globulins. As found in seeds of diverse dicotyledonous species, several 2S albumin isoforms are present in sesame seed. In comparison with the extensive studies on sesame 11S globulins, not much investigation has been focused on sesame 2S albumin until it was identified as a severe allergen in the past two decades (Beyer et al. 2002).

3.2.4 A Minor Storage Protein: 7S Globulin

Three classes of storage proteins, 11S globulin, 7S globulin, and 2S albumin, are commonly found in dicotyledonous seeds, and isoforms are usually detected in each of the three classes (Shewry 1995). Similarly, besides the abundant 11S globulin and 2S albumin isoforms, two 7S globulin isoforms of 55–60 kDa are detected as minor constituents in protein bodies of sesame seed and comprise approximately 1–2% of the total sesame seed proteins (Tai et al. 2001). Different from 11S globulin and 2S albumin, the mature polypeptide of 7S globulin is composed of a single subunit with no consensus cleavage site for the splitting protease, asparaginyl endopeptidase (Fig. 3.2). Mature 7S globulin isoforms exist in both soluble and insoluble fractions of sesame seed protein extract.

3.2.5 Accumulation of Storage Proteins in Protein Bodies

Many protein bodies (approximately 5 μm in diameter) containing a matrix of storage proteins surrounded by a lipid bilayer are observed in the cytosol of mature sesame seed cells (Fig. 3.1). Immunolocalization reveals the coexistence of three classes of storage proteins, 11S globulin, 7S globulin, and 2S albumin, in each of the

protein bodies (Tai et al. 2001). While 2S albumin and 7S globulin are mostly detected in the peripheral region of the protein body, 11S globulin is evenly detected in the whole protein body. Generally, the deposition of storage proteins occurs in ER membrane of maturing seeds where they are processed, modified, and assembled to form protein bodies before being transported to the cytosol. An N-terminal cleavable signal peptide essential for ER targeting via signal recognition particle (SRP)-dependent pathway is present in the precursor polypeptides of these three types of storage proteins. After ER penetration, 11S globulin and 2S albumin, but not 7S globulin, are processed into two subunits after forming intramolecular disulfide bonds (Hara-Nishimura et al. 1998). The cleavage is possibly processed by the same protease, i.e., asparaginyl endopeptidase. A cNDA fragment (accession number AF169973) encoding the precursor polypeptide of a putative asparaginyl endopeptidase has been cloned from maturing sesame seeds. After proteolytic cleavage, the three classes of sesame storage proteins are processed, modified, assembled, and deposited into vacuoles of maturing protein bodies, presumably in a compatible route of transportation.

3.2.6 Gene Families Encoding Sesame Storage Proteins

Isoforms of sesame 11S globulin, 7S globulin, or 2S albumin, instead of being posttranslationally modified from single gene products, are presumably encoded by gene families. To date, cDNA sequences encoding precursor polypeptides of four 11S globulins (Accession numbers: AF091842, AF240004, DQ256293, and DQ256294), three 2S albumins (Accession numbers: AF091841, AF240005 and DQ256292), and one 7S globulin (Accession number: AF240006) have been cloned from maturing sesame seeds. Alignments of the deduced amino acid sequences of the precursor polypeptides of sesame 11S globulins and 2S albumins are separately shown in Figs. 3.3 and 3.4.

A database of 3328 expressed sequence tag (EST) sequences from maturing sesame seeds was established (Suh et al. 2003). Among these sequences, 44, 9, 19, and 28 EST sequences are corresponding to cDNA fragments encoding the four 11S globulin isoforms, respectively; 25, 46, and 13 EST sequences are corresponding to cDNA fragments encoding the three 2S albumin isoforms, separately. However, no EST sequences corresponding to cDNA fragments encoding the 7S globulin isoforms are found in this database. Possibly, the expression levels of 7S genes are relatively low during seed maturation.

All genes encoding the precursor polypeptides of 11S globulins, 7S globulins, and 2S albumins are presumably present as single or low copy number in sesame genome; and they are assumed to be specifically turned on during seed maturation when protein bodies are actively assembled (Tai et al. 2001). All mRNAs of the three classes of sesame storage proteins start to accumulate in maturing seeds approximately two weeks after flowering, diminish in the seed-drying stage, and eventually vanish in mature seeds. In detailed analyses of maturing sesame seeds by Northern and Western blotting analyses, the expression and accumulation of 7S globulin occur slightly lagging behind those of 11S globulin and 2S albumin.

3.2.7 Deduced Amino Acid Compositions of Sesame Storage Proteins

For a better understanding of sesame nutritive protein composition, amino acid compositions of sesame storage proteins were deduced from the available cDNA sequences (Table 3.2). As expected, some of these sesame abundant proteins including 2S-1, 2S-3, and 11S-1 do comprise relatively high methionine and cysteine contents and thus account for the nutritive value of sesame proteins. Particularly, a quarter of amino acids in 2S-1 are sulfur-containing amino acids; the contents of methionine and cysteine are 13.64 and 9.09%, respectively. In a word, the

Fig. 3.3 Sequence alignment of the deduced precursor polypeptides of the four 11S globulin isoforms in sesame seeds. The amino acid number for the last residue in each row is listed on the right for each isoform. Broken lines in the sequences represent gaps introduced for best alignment and conserved residues are shaded

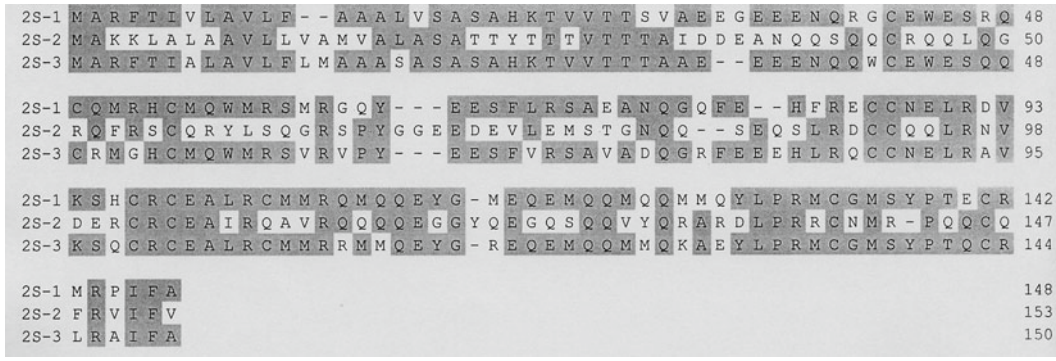


Fig. 3.4 Sequence alignment of the deduced precursor polypeptides of the three 2S albumin isoforms in sesame seeds. The amino acid number for the last residue in each row is listed on the right for each isoform. Broken lines in the sequences represent gaps introduced for best alignment and conserved residues are shaded

Table 3.2 Amino acid composition of different storage proteins in sesame seeds. Methionine and cysteine contents (% by frequency) are calculated according to the mature protein sequences deduced from their corresponding genes

Isoform	11S-1	11S-2	11S-3	11S-4	2S-1	2S-2	2S-3	7S
A (Ala)	6.62	7.16	6.84	8.91	3.64	2.61	6.19	5.86
C (Cys)	1.14	0.84	0.85	1.34	9.09	6.96	8.85	2.31
D (Asp)	4.11	5.26	2.78	3.79	0.91	3.48	0.88	3.37
E (Glu)	7.53	7.16	9.19	7.35	11.82	7.83	11.50	11.90
F (Phe)	3.65	4.21	3.63	4.45	3.64	2.61	2.65	4.26
G (Gly)	7.53	7.79	9.62	9.13	4.55	6.96	3.54	8.88
H (His)	2.51	1.68	2.14	2.45	2.73	0.00	1.77	2.49
I (Ile)	4.34	4.00	4.70	5.35	0.91	1.74	0.88	4.44
K (Lys)	1.83	2.53	1.92	1.34	0.91	0.00	1.77	4.44
L (Leu)	6.62	9.05	7.69	6.24	3.64	5.22	4.42	5.68
M (Met)	3.20	1.68	0.85	1.56	13.64	1.74	10.62	1.42
N (Asn)	4.79	4.84	5.56	5.79	1.82	2.61	1.77	4.09
P (Pro)	4.34	4.63	3.63	4.01	2.73	2.61	2.65	3.55
Q (Gln)	8.68	7.16	8.97	8.46	14.55	24.35	12.39	9.59
R (Arg)	9.82	10.32	10.47	10.02	12.73	14.78	13.27	10.83
S (Ser)	7.53	5.47	5.98	6.68	5.45	6.96	5.31	5.33
T (Thr)	5.25	4.84	5.56	2.67	0.91	0.87	0.88	3.37
V (Val)	6.85	6.53	5.77	6.46	0.91	5.22	4.42	4.97
W (Trp)	0.91	1.68	0.85	1.11	1.82	0.00	2.65	0.53
Y (Tyr)	2.74	3.16	2.99	2.90	3.64	3.48	3.54	2.66

sesame richness in sulfur-containing amino acids is mainly attributed to the soluble 2S albumin.

3.3 Sesame Oil

Oil content of sesame varies significantly (ranging from 35 to 65% of dry weight), depending on varieties and their cultivation conditions with the average oil content of approximately 50% (Namiki 1995). Sesame oil is composed of triacylglycerols (>95%), diacylglycerols, free fatty acids, phospholipids, and some unsaponifiables. Similar to many other seed oils, sesame oil contains an equimixture of oleic and linoleic acids. Sesame oil contains roughly 500 mg/kg total tocopherols, including abundant γ-tocopherol (95.5–99%), minor δ-tocopherol, and trace amounts of α-tocopherol and β-tocopherol (Kamal-Eldin and Appelqvist 1994). However, the total tocopherol level in sesame oil is generally lower than that in other vegetable oils (Kamal-Eldin 2005). Empirically, sesame oil has long been used as health food in Asian countries. In comparison with most other edible oils extracted from diverse seeds, sesame oil is extremely stable partly due to the effective antioxidant activities attributed to its abundance of oil-soluble furofuran lignans, mainly sesamin and sesamolin (Fig. 3.5). The contents of sesamin and sesamolin in sesame oils vary significantly among species; the content of sesamin ranges from 700 to 1130 mg/kg while that of sesamolin ranges from 200 to 590 mg/kg (Tashiro et al. 1990).

3.3.1 Sesamin and Sesamolin

Sesamin and sesamolin were found to possess several biological functions including suppressing lipid peroxidation in erythrocytes, inhibiting intestinal absorption of cholesterol and hepatic 3-hydroxy-3-methylglutaryl CoA reductase activity, preventing chemically induced mammary cancer, inhibiting Δ^5-desaturase and chain elongation of C18 fatty acids, exhibiting the antihypertensive effect, enhancing liver detoxification of carbon tetrachloride and ethanol, and quenching the excess generation of nitric oxide (NO) induced by lipopolysaccharide (LPS) in rat primary microglia cells (Dar and Arumugam 2013; Hou et al. 2003). Moreover, sesamin and sesamolin exert effective neuroprotection against cerbral ischemia as observed in vivo using gerbils subjected to a focal cerebral ischemia induced by occlusion of the right common carotid artery and the right middle cerebral artery (Cheng et al. 2006). These two furofuran lignans were also demonstrated to significantly alleviate liver damage of rats caused by carbon tetrachloride in combination with kava (Chen et al. 2010).

3.3.2 Sesame Enterolignan Precursors

Enterolignans (also called mammalian lignans), such as enterodiol and enterolactone, are generated by the human intestinal microflora after the consumption of certain plant lignans (Coulman

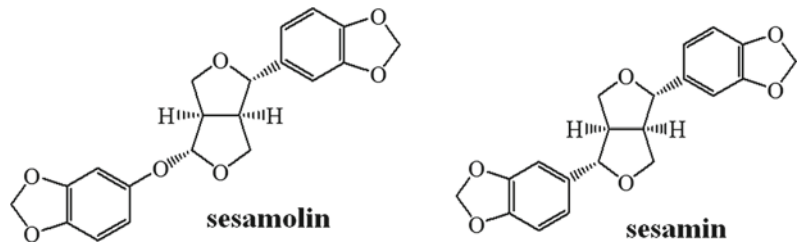


Fig. 3.5 Chemical structures of sesamolin and sesamin

et al. 2005). Due to their structural similarity to estrogen and their capability of binding to estrogen receptors, enterolignans are proposed to be weak estrogenic/antiestrogenic analogs. Therefore, the consumption of enterolignans or their precursor molecules is suggested to reduce the risk of hormone-related diseases as well as certain cancers and coronary heart diseases. It has been noticed that sesamin and sesamolin from sesame oil are adequate precursors from food sources for the generation of enterolignans in human body (Peñalvo et al. 2005; Liu et al. 2006; Wu et al. 2006).

3.4 Sesame Oil Bodies

The high-quality oil molecules, including triacylglycerols and furofuran lignans, in sesame seed are stored in specialized organelles termed oil bodies of average sizes approximately 2 μm (Fig. 3.1). An oil body in sesame seed comprises an oil matrix that is surrounded by a monolayer of phospholipids embedded with proteins (Fig. 3.6), and the proteins covering almost the entire surface of oil bodies sustain their integrity and stability by providing steric hindrance and electronegative repulsion among these oil storage organelles (Tzen 2012). Therefore, the compacted oil bodies in mature sesame seed cells by no means fuse or combine, even the seeds have been stored for a period of time at room temperature. The three essential constituents of sesame oil bodies, namely matrix oil, phospholipid, and protein are 98.5, 0.6, and 0.7% by weight, respectively; and the phospholipids found in sesame oil bodies are 41.2% phosphatidylcholine, 15.8% phosphatidy lethanolamine, 20.9% phosphatidylinositol, and 22.1% phosphatidylserine (Tzen et al. 1993).

3.4.1 Historical Studies on Integral Proteins of Sesame Oil Bodies

Integral proteins on the surface of sesame oil bodies have been actively investigated by using tools of molecular biology and protein chemistry in the past three decades. Three classes of oil body-associated proteins, oleosin, caleosin, and steroleosin, have been identified in sesame seeds (Fig. 3.7), and isoforms are present in each of the three classes (Chen et al. 1997, 1998, 1999; Frandsen et al. 2001; Lin et al. 2002, 2005; Tai et al. 2002; Tzen et al. 1997). Oleosin, the most abundant protein class on the surface of sesame oil bodies presumably serves as the major structural component to maintain the physical and chemical organization of these lipid storage organelles (Tzen et al. 2003). This abundant protein is characterized by a distinctive central hydrophobic domain of approximately 8 kDa (around 70 non-hydrophilic amino acid residues). Caleosin has a calcium-binding motif and several potential phosphorylation sites that are proposed to be modulated for some physiological roles in the formation or degradation of oil bodies (Chen and Tzen 2001). The central hydrophobic domain putatively responsible for anchoring caleosin to oil bodies is around 4 kDa. Steroleosin is assumed to contain a sterol-binding dehydrogenase belonging to a super-family of pre-signal proteins involved in signal transduction controlled by their binding sterols (Lin and Tzen 2004). Different from the central hydrophobic domains of oleosin and caleosin, steroleosin is assumed to anchor to oil bodies by its N-terminal domain of approximately 5 kDa. These three classes of oil body proteins found in sesame are presumably present in seed oil bodies of diverse angiosperm species. Moreover, the structural organization of sesame oil bodies is applicable to oil bodies in gymnosperm species (Wu et al. 1999; Jiang et al. 2009; Pasaribu et al. 2014, 2016, 2017).

3.4.2 Gene Families Encoding Sesame Oil Body Proteins

All mRNAs coding for sesame oleosin, caleosin, and steroleosin isoforms specifically accumulate in maturing seeds when oil bodies are actively assembled, and subsequently diminish when seeds become mature (Chen et al. 1999; Lin and

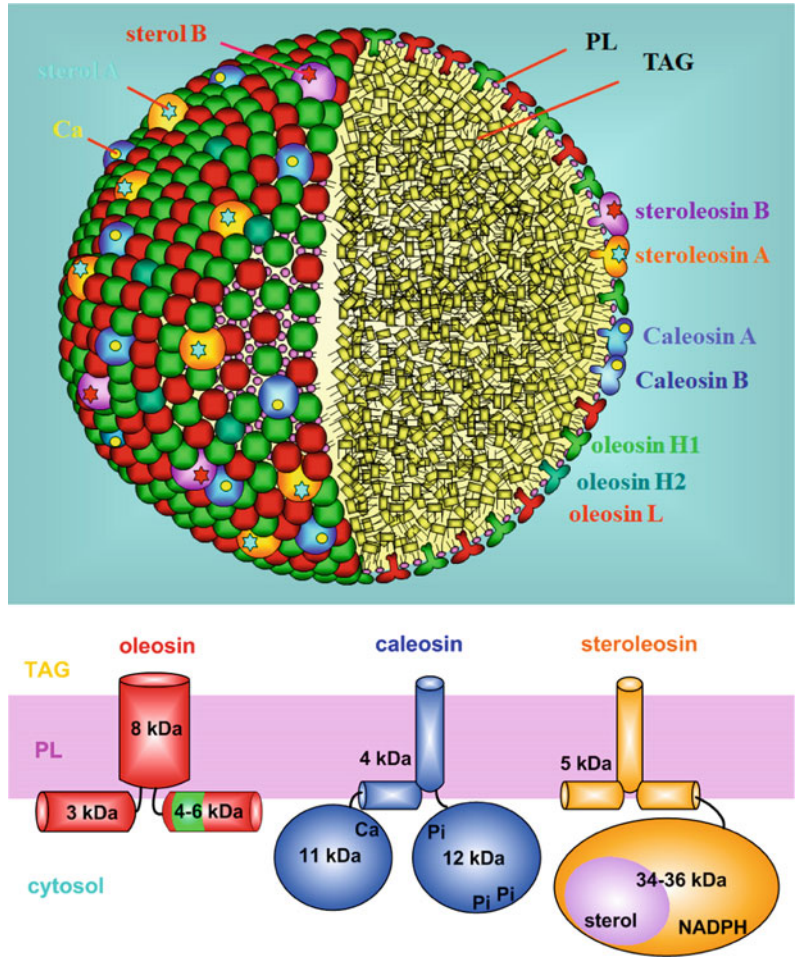


Fig. 3.6 Model of a sesame oil body. Simplified structures of oleosin, caleosin, and steroleosin are shown at the bottom. Modified from Lin and Tzen (2004)

Tzen 2004; Tai et al. 2002). So far, cDNA sequences corresponding to three oleosin isoforms (Accession numbers: AF302807, U97700, and AF091840), two caleosin isoforms (Accession numbers: AF109921 and DQ088381), and two steroleosin isoforms (Accession numbers: AF302806 and AF498264) have been cloned from maturing sesame seeds. Deduced amino acid sequences of these three classes of oil body proteins and their homologous structural regions revealed in sequence alignments are separately shown in Figs. 3.8, 3.9, and 3.10.

In a database search of 3328 EST sequences from maturing sesame seeds (Suh et al. 2003), EST sequences corresponding to abundant oleosin-H1 and oleosin-L isoforms could be identified. However, no EST sequences corresponding to the minor oleosin-H2, caleosin isoforms, and steroleosin isoforms were identified in these sesame EST sequences, presumably due to their low expression levels in the maturing sesame seeds. Apparently, a random sequencing of around 3000 EST sequences seems to be insufficient for functional genomics analysis of minor oil body proteins in sesame seeds.

3.4.3 Isoforms of Sesame Oil Body Proteins

To date, all known seed oleosins can be classified as one of the two isoforms, i.e., H- and L- (high and low molecular weight) oleosins (Tzen et al. 1990). Two H-oleosins (17 and 15.5 kDa) and

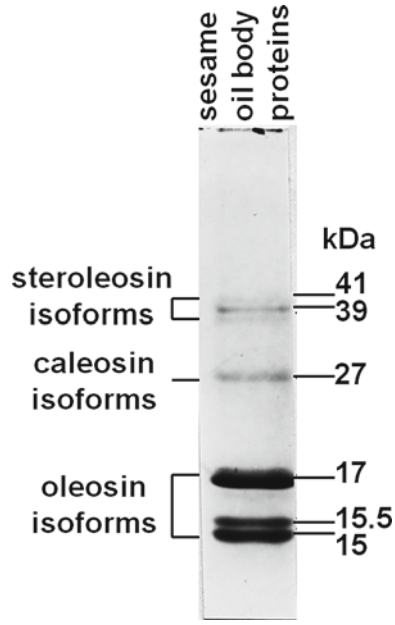


Fig. 3.7 SDS-PAGE of total proteins in sesame seed oil bodies

one L-oleosin (15 kDa) are detected in sesame oil bodies. The physiological significance of the presence of these three oleosin isoforms remains to be elucidated. Two caleosin isoforms of similar molecular mass (27 kDa) are present in sesame oil bodies. While oleosin isoforms are exclusively found in oil bodies, caleosin isoforms are detected in other cellular fractions, e.g., endoplasmatic reticulum (Næsted et al. 2000). Whether sesame caleosin isoforms are also present in other subcellular fractions besides seed oil bodies remains to be confirmed. Two steroleosin isoforms of 41 and 39 kDa are observed in sesame seed oil bodies. Similar to caleosin isoforms, steroleosin isoforms are possibly located in other cellular fractions besides seed oil bodies (Lin and Tzen 2004). The apparent distinction of the two steroleosin isoforms in sesame oil bodies is mainly resulted from their distinct sterol-binding sub-domains, and this distinction implies that the two sesame steroleosin isoforms may be modulated by different sterols to regulate two distinct physiological roles for the formation or degradation of seed oil bodies. However, the endogenous ligands (sterols) for the two steroleosin isoforms in maturing or germinating sesame seeds have not been identified so far.

3.4.4 Targeting and Assembly of Sesame Oil Body Proteins

How oleosin, caleosin, and steroleosin are assembled to sesame oil bodies has not been well illustrated. To initiate this study, an in vitro system was established; artificial oil emulsions were constituted to mimic maturing seed oil bodies for the simultaneously targeting of in vitro translated oil body proteins (Chen and Tzen 2001). The results suggest that calesoin or oleosin spontaneously targets to maturing oil bodies

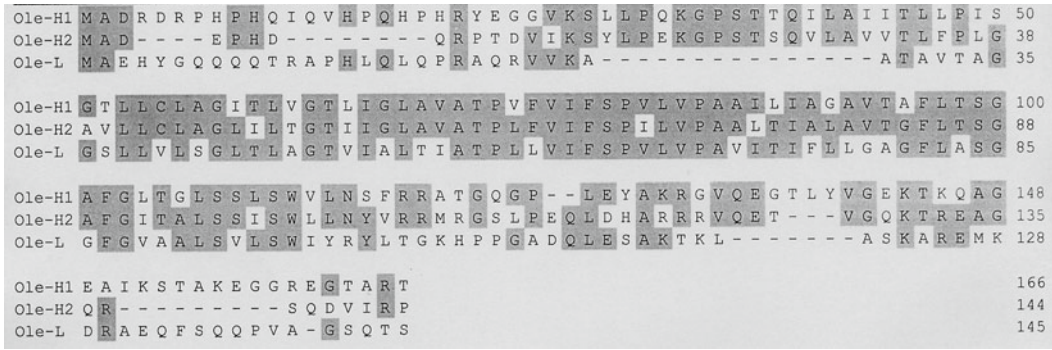


Fig. 3.8 Sequence alignment of the three deduced oleosin isoforms in sesame seeds. The amino acid number for the last residue in each row is listed on the right for each isoform. Broken lines in the sequences represent gaps introduced for best alignment, and conserved residues are shaded

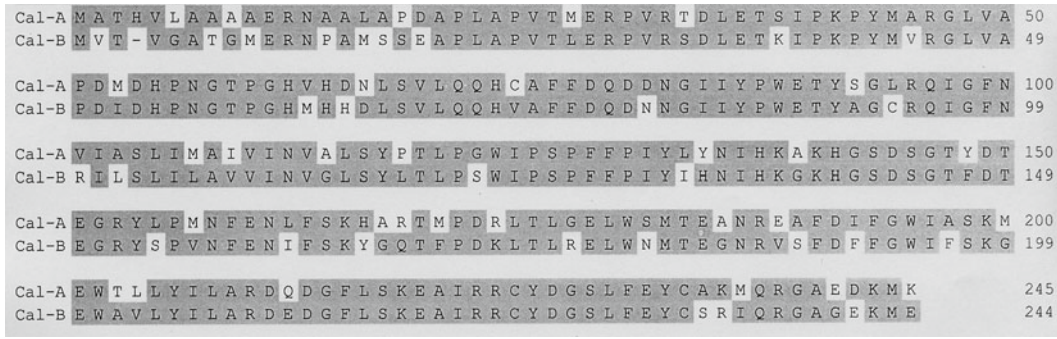


Fig. 3.9 Sequence alignment of the two deduced caleosin isoforms in sesame seeds. The amino acid number for the last residue in each row is listed on the right for each isoform. Broken lines in the sequences represent gaps introduced for best alignment, and conserved residues are shaded

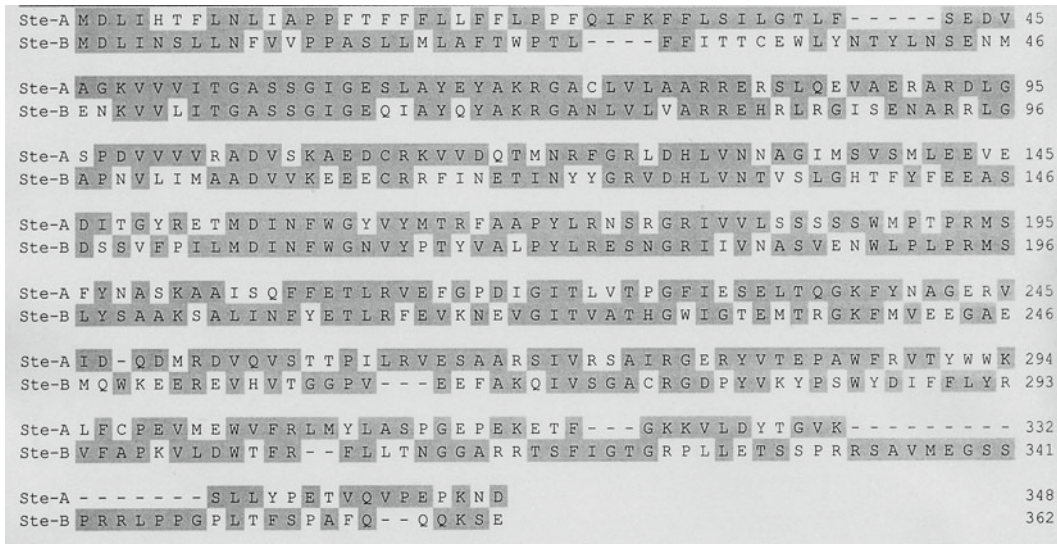


Fig. 3.10 Sequence alignment of the two deduced steroleosin isoforms in sesame seeds. The amino acid number for the last residue in each row is listed on the right for each isoform. Broken lines in the sequences represent gaps introduced for best alignment, and conserved residues are shaded

where ER membranes are concurrently expanded with deposited lipid molecules to form an oil matrix. In contrast, steroleosin is assembled to maturing oil bodies through different locations of endoplasmic reticulum (ER) membrane (Tzen et al. 2003; Tzen 2012). It seems that steroleosin is directed by its hydrophobic N-terminal domain, which serves as an uncleavable signal peptide for ER targeting via SRP-dependent pathway. Subsequently, steroleosin is translocated to the phospholipid bilayer of ER membrane and then laterally transported to the surface of maturing sesame oil bodies (Fig. 3.11).

3.5 Artificial Sesame Oil Bodies

With proper proportions of the three essential constituents in native sesame oil bodies, namely matrix oil, membrane phospholipid, and surfactant protein, artificial sesame oil bodies of similar sizes

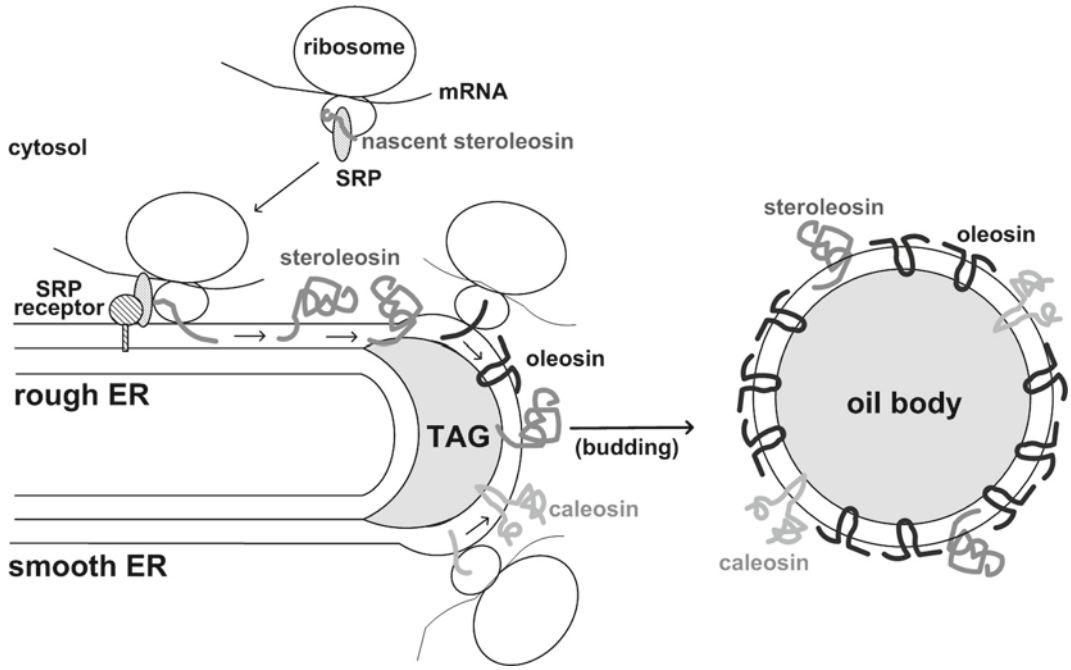


Fig. 3.11 Targeting model of oleosin, caleosin, and steroleosin to a maturing sesame oil body. Modified from Tzen (2012)

(0.5–2 μm) and structural stability were successfully constituted by sonication (Tai et al. 2002). Stable artificial sesame oil bodies were generated by the combination of oleosin isoforms or any oleosin isoform alone, i.e., oleosin-H1, oleosin-H2, or oleosin-L; relatively, oleosin-L provided slightly better structural stability than the other two isoforms. Recombinant oleosin expressed in *Escherichia coli* was also able to stabilize artificial sesame oil bodies with size, topology and stability comparable to those encapsulated with native oleosin isolated from sesame seeds (Tzen et al. 1998). Furthermore, sizes of artificial sesame oil bodies could be controlled by changing the ratio of matrix oil over oil body protein (mainly oleosin isoforms), and a normal distribution with an average size proportional to this ratio was displayed (Peng et al. 2003). Thermostability of artificial sesame oil bodies was found to be successively reduced as their sizes were subsequently increased. When surfactant proteins (mainly oleosin isoforms) were cross-linked by glutaraldehyde or genipin, the thermally tolerant temperature of artificial sesame oil bodies could be uplifted from 50 to 90 °C.

Artificial sesame oil bodies of approximately 10 times smaller (50–200 nm) than native sesame oil bodies were successfully constituted with matrix oil, membrane phospholipid, and caleosin (Chen et al. 2004). These relatively small artificial sesame oil bodies constituted with caleosin possessed higher thermostability (up to 70 °C), and they were consistently kept as discrete particles at lower pH surroundings, in comparison with those stabilized by oleosin. Recombinant sesame caleosin expressed in *E. coli* could stabilize artificial sesame oil bodies with size, topology, and stability comparable to those encapsulated with native caleosin isolated from sesame seeds. Under the same reconstitution condition, artificial sesame oil bodies were unstable and decomposed rapidly when oleosin or caleosin was replaced with equivalent quantity of steroleosin.

3.5.1 A Protein Expression/Purification System

A bacterial expression/purification system has been developed to produce recombinant proteins efficiently by massively accumulating target proteins fused to oleosins on the surface of artificial sesame oil bodies (Peng et al. 2004a, b). In this system, a target protein was first over-expressed as an insoluble oleosin-fused recombinant protein largely aggregated in the pellet of cell lysate. The oleosin-fused target protein was simply harvested by centrifugation, assembled into artificial sesame oil bodies, released via a specific proteolytic cleavage, and then collected by concentrating the ultimate supernatant. For the purpose of protein purification, this new technique is a powerful and competitive method in comparison with several protocols using conventional affinity chromatography. However, it is also relatively expensive for the enzymatic release of the target protein from the recombinant oleosin-fused polypeptide on the surface of artificial sesame oil bodies. To solve this economical problem, an improved system has been developed and demonstrated by producing nattokinase in *E. coli* (Chiang et al. 2005). In this revised system, nattokinase was first over-expressed as an insoluble oleosin-intein-fused protein in the pellet of cell lysate. The oleosin-intein-nattokinase was harvested by centrifugation, assembled into artificial sesame oil bodies, released via self-splicing of the intein linker, and then collected by concentrating the ultimate supernatant (Fig. 3.12). Without using an expensive protease for specific cleavage, the release of target protein via intein self-splicing induced by temperature adjustment significantly reduces the processing cost. Apparently, this improved system provides an inexpensive process to produce recombinant proteins with high efficiency.

In a different viewpoint, the target protein fused to oleosin may be used without releasing from artificial oil bodies; a novel system of enzyme fixation was designed to achieve, in one step, protein refolding and immobilization by linking an active enzyme. For example, D-hydantoinase, fused to sesame oleosin on the surface of artificial oil bodies, has been engineered and used directly (Chiang et al. 2006). The immobilized D-hydantoinase remained as an active enzyme for many cycles of industrial reaction and kept its activity in solution after stored at 4 °C for two weeks. Apparently, this is a useful system for enzyme immobilization.

3.5.2 A Carrier for Drug Delivery

Oral delivery is still the most important and frequently used application route for drug administration. However, oral bioavailability of many hydrophobic drugs is extremely low and thus evidently reduced their efficacy in medicinal utilization. In contrast with the regular sizes of artificial sesame oil bodies stabilized with oleosins, the relatively small artificial sesame oil bodies, constituted with matrix oil, phospholipid, and caleosin, have been used to develop an oral delivery system for hydrophobic drugs as exemplified by the delivery of cyclosporine A, a hydrophobic drug used in the treatment of some autoimmune illnesses (Chen et al. 2005). In the preparation of delivery formulation, cyclosporine A was successfully trapped in artificial sesame oil bodies, and the drug-loaded artificial oil bodies remained stable after stored at 4 °C for weeks. This designed delivery formulation showed a satisfactory bioavailability in an animal study via oral administration.

Though curcumin is a hydrophobic molecule, it is found to be indissoluble in most seed oils. Thus, a mixed dissolvent formula was designed to dissolve curcumin. Furthermore, artificial oil bodies constituted with a recombinant sesame caleosin were employed to encapsulate the admixture containing curcumin (Chang et al. 2013a, b). These artificial oil bodies of relatively small sizes (150 nm) were successfully solidified in the forms of powder and tablet (Fig. 3.13). Oral bioavailability of curcumin was demonstrated to be significantly elevated via encapsulation in this novel formulation of artificial oil bodies in an animal study. Artificial oil bodies

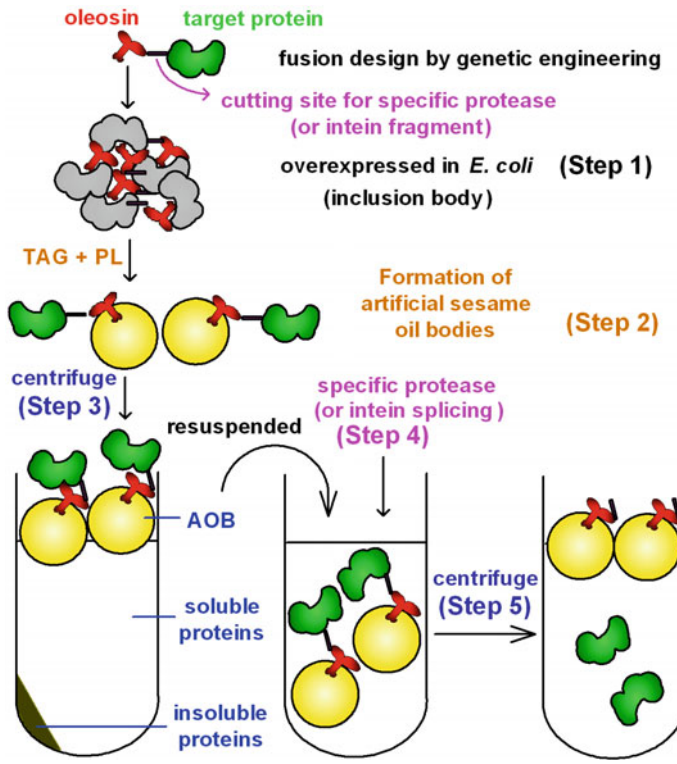


Fig. 3.12 Scheme of the bacterial expression/purification system via artificial sesame oil bodies (modified from Peng et al. 2004a). Step 1: overexpression of a target protein in an insoluble oleosin-fused form (gray color representing an incorrect folding) in the inclusion body of *E. coli*. Step 2: reconstitution of artificial sesame oil bodies with triacylglycerol (TAG), phospholipid (PL), and the insoluble oleosin-fused protein. Step 3: collection of ASOB by centrifugation. Step 4: separation of the target protein from oleosin in artificial sesame oil bodies via proteolytic cleavage or self-splicing of intein. Step 5: harvesting the target protein from the ultimate supernatant after centrifugation

seem to provide an adequate formulation to be evaluated for the possibility of encapsulating many other hydrophobic drugs for oral administration.

3.5.3 A Carrier to Render Haptens for Antibody Production

Immunodetection based on the specific interaction between antigens and antibodies has been demonstrated to be a powerful tool in pharmaceutical diagnosis as well as in academic studies. Enzyme-linked immunosorbent assay (ELISA) is commonly used to quickly screen target molecules, such as marker proteins of diseases (Butler 2000). However, relatively tiny molecules, i.e., less than 1000 Da, are not immunogenic and hardly elicit immune responses in animals. Therefore, to generate antibodies against a small molecule, a carrier protein is typically employed to conjugate with the tiny molecule and form an artificial antigen (protein molecule complex) prior to animal immunization. Frequently, antibodies against the tiny molecule are weakly generated while antibodies against the carrier protein are massively generated. To establish a suitable system to render tiny molecules for antibody production, recombinant sesame caleosin was engineered to possess 17 extra Lys residues, overexpressed in *E. coli*, chemically linked to target tiny molecules, and assembled to artificial sesame oil bodies prior to animal immunization. (Liu et al. 2011). By using this system, mono-specific

Fig. 3.13 Preparation of artificial oil bodies encapsulating curcumin. Modified from Chang et al. (2013a)

polyclonal antibodies against biotin were successfully generated in rats (Fig. 3.14). Interestingly, those antibodies generated via the biotinylated caleosin on the surface of artificial sesame oil bodies no longer recognized sesame caleosin. It appears that the engineered Lys-rich caleosin is a suitable carrier protein for the production of antibodies against tiny molecules via presentation of target compounds on the surface of artificial sesame oil bodies.

3.5.4 A Detecting Kit for Tea Astringency

Astringency is a mouth sensation of tea or fruit due to the massive aggregation of polyphenolic compounds and salivary proteins, and the complex aggregation significantly reduces lubricating function of saliva (Bajec and Pickering 2008). A practical assay was designed to quantitatively evaluate the relative astringency of oolong tea in vitro by using artificial sesame oil bodies sheltered by a modified caleosin fused with histatin 3, a kind of human salivary peptide (Shih et al. 2017). After mixed with oolong tea infusion, the artificial oil bodies tend to trap tea polyphenol and float on top of the solution to form a visible milky layer. The thickness of aggregated artificial oil bodies seems to be corresponding to the tea astringency (Fig. 3.15). Relative astringency of tea infusions can be calculated by recording the visible thickness of aggregated artificial oil bodies in comparison with the standard curve generated by concurrently observing a serial dilution of standard tea infusions. It seems to be an adequate kit to detect relative astringency of tea and other drinks.

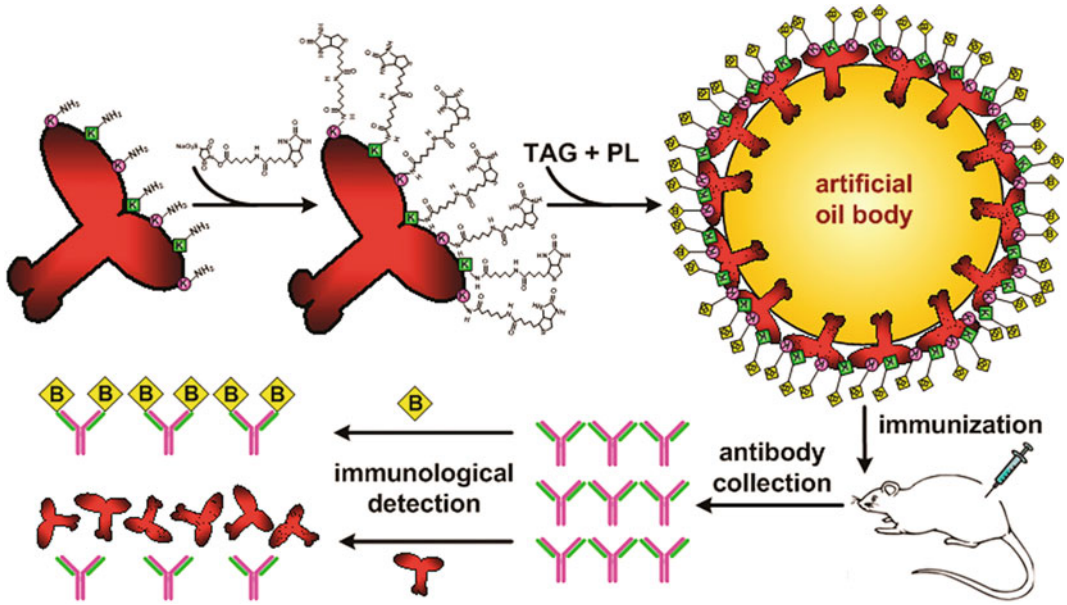


Fig. 3.14 Scheme of producing antibodies against biotin via artificial sesame oil bodies constituted with biotinylated caleosins. Modified from Liu et al. (2011)

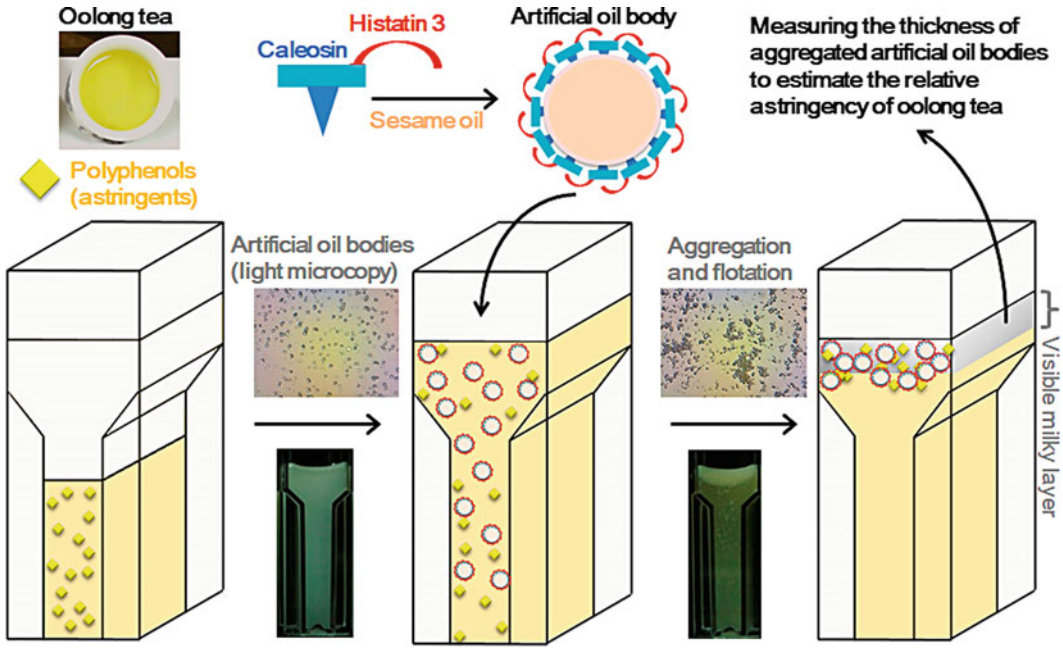


Fig. 3.15 Scheme of detecting tea astringency via observing flotation of artificial oil bodies constituted with caleosin fused with a salivary protein, histatin 3. Aggregated artificial sesame oil bodies that floated and formed milky layers on top of sample solutions in cuvettes were photographed at different time intervals after mixed with different tea infusions. Relative astringency of tea infusions was correlated to the thickness of floated artificial oil bodies. Modified from Shih et al. (2017)

3.6 Concluding Remarks

Sesame is regarded as a nutritive protein source due to its richness in sulfur-containing amino acids, which is mainly attributed to the soluble 2S albumin. Sesame oil, conventionally regarded as a high-quality edible oil source in Asia, is highly attributed to its furofuran lignans, mainly sesamin and sesamolin. In the past three decades, structure and organization of sesame oil bodies have been well characterized. Oleosin, caleosin, and steroleosin, the three classes of proteins identified in sesame oil body, have served as references for the studies on the integral proteins of lipid storage organelles in other plant species. Gene families encoding these three groups of oil body proteins have been sequenced nearly completely from maturing sesame seeds. Oleosin- and caleosin-stabilized artificial oil bodies are successfully constituted and used to develop a bacterial expression/purification system, a hydrophobic drug delivery system, an antibody production system, and a visible detecting kit for tea astringency. Further investigation and technical improvement are expected to create novel artificial sesame oil bodies as versatile vehicles that could fulfill many other requirements for specialized applications.

Acknowledgements The research work on sesame was continually supported by grants from the Ministry of Science and Technology, Taiwan, for over two decades. The author cordially expresses appreciation to all the colleagues and graduate students for their synergistic contribution.

References

Bajec MR, Pickering GJ (2008) Astringency: mechanisms and perception. Crit Rev Food Sci 48:858–875

Beyer K, Bardina L, Grishina G, Sampson HA (2002) Identification of sesame seed allergens by 2-dimensional proteomics and Edman sequencing: seed storage proteins as common food allergens. J Allergy Clin Immunol 110:154–159

Butler JE (2000) Enzyme-linked immunosorbent assay. J Immunoassay 21:165–209

Chang MT, Tsai TR, Lee CY, Wei YS, Chen YJ, Chen CR, Tzen JTC (2013a) Elevating bioavailability of curcumin via encapsulation with a novel formulation of artificial oil bodies. J Agri Food Chem 61:9666–9671

Chang MT, Chen CR, Liu TH, Lee CP, Tzen JTC (2013b) Development of a protocol to solidify native and artificial oil bodies for long-term storage at room temperature. J Sci Food Agri 93:1516–1519

Chen JCF, Tzen JTC (2001) An in vitro system to examine the effective phospholipids and structural domain for protein targeting to seed oil bodies. Plant Cell Physiol 42:1245–1252

Chen JCF, Lin RH, Huang HC, Tzen JTC (1997) Cloning, expression and isoform classification of a minor oleosin in sesame oil bodies. J Biochem 122:819–824

Chen ECF, Tai SSK, Peng CC, Tzen JTC (1998) Identification of three novel unique proteins in seed oil bodies of sesame. Plant Cell Physiol 39:935–941

Chen JCF, Tsai CCY, Tzen JTC (1999) Cloning and secondary structure analysis of caleosin, a unique calcium-binding protein in oil bodies of plant seeds. Plant Cell Physiol 40:1079–1086

Chen MCM, Chyan CL, Lee TTT, Huang SH, Tzen JTC (2004) Constitution of stable artificial oil bodies with triacylglycerol, phospholipid and caleosin. J Agri Food Chem 52:3982–3987

Chen MCM, Wang JL, Tzen JTC (2005) Elevating bioavailability of cyclosporine A via encapsulation in artificial oil bodies stabilized by caleosin. Biotechnol Prog 21:1297–1301

Chen WL, Lu HC, Huang HY, Hwang GY, Tzen JTC (2010) Sesame lignans significantly alleviate liver damage of rats caused by carbon tetrachloride in combination with kava. J Food Drug Anal 18:225–231

Cheng FC, Jinn TR, Hou RCW, Tzen JTC (2006) Neuroprotective effects of sesamin and sesamolin on gerbil brain in cerebral ischemia. Int J Biomed Sci 2:284–288

Chiang CJ, Chen HC, Chao YP, Tzen JTC (2005) Efficient system of artificial oil bodies for functional expression and purification of recombinant nattokinase in *E. coli*. J Agri Food Chem 53:4799–4804

Chiang CJ, Chen HC, Kuo HF, Chao YP, Tzen JTC (2006) A simple and effective method to prepare immobilized enzymes using artificial oil bodies. Enzyme Microb Technol 39:1152–1158

Coulman KD, Liu Z, Hum WQ, Michaelides J, Thompson LU (2005) Whole sesame seed is as rich a source of mammalian lignan precursors as whole flaxseed. Nutr Cancer 52:156–165

Dar AA, Arumugam N (2013) Lignans of sesame: purification methods, biological activities and biosynthesis—a review. Bioorg Chem 50:1–10

Frandsen GI, Mundy J, Tzen JTC (2001) Oil bodies and their associated proteins, oleosin and caleosin. Physiol Plant 112:301–307

Fujino S, Kamiguchi Y, Yokoi Y, Hasegawa K (1981) Effects of reductive alkylation on gelation of sesame 13S globulin. Nippon Nogei Kagaku Kaishi 55:975–981

Hara-Nishimura I, Shimada T, Hatano K, Takeuchi Y, Nishimura M (1998) Transport of storage proteins to

protein storage vacuoles is mediated by large precursor-accumulating vesicles. Plant Cell 10:825–836

Hasegawa K, Murata M, Fujino S (1978) Characterization of subunits and temperature-dependent dissociation of 13S globulin of sesame seed. Agric Biol Chem 42:2291–2297

Hasegawa K, Owaki K, Tanaka T, Fujino S (1979) Gelation characteristics and turbidities of sesame protein and soybean proteins. Nippon Nogei Kagaku Kaishi 53:375–384

Hasegawa K, Tanaka T, Tamai S (1981) Preparation of hybrid proteins from subunits of sesame 13S and soybean 11S globulins and their molecular weights. Agri Biol Chem 45:809–815

Hou RCW, Chen HL, Tzen JTC, Jeng KCG (2003) Protective effects of sesamin and sesamolin on hypoxic neuronal and PC12 cells. J Neurosci Res 74:123–133

Hsiao ESL, Lin LJ, Li FY, Wang MMC, Liao MY, Tzen JTC (2006) Gene families encoding isoforms of two major sesame seed storage proteins, 11S globulin and 2S albumin. J Agri Food Chem 54:9544–9550

Jiang PL, Chen JCF, Chiu ST, Tzen JTC (2009) Stable oil bodies sheltered by a unique caleosin in cycad megagametophytes. Plant Physiol Biochem 47:1009–1016

Jones DB, Gersdorff CEF (1927) Proteins of sesame seed, *Sesamum Idicum*. J Biol Chem 75:213

Kamal-Eldin A (2005) Minor components in vegetable oils, chapter 12. In: Shahidi F (ed) Baileys industrial fats and oils, vol 3, edible oil and fat products: specialty oils and oil products. Wiley, Sussex, UK

Kamal-Eldin A, Appelqvist LÅ (1994) Variations in the composition of sterols, tocopherols and lignans in seed oils from four *Sesamum* species. J Am Oil Chem Soc 71:149–156

Lakashmi TS, Nandi PK (1979) Studies on the effect of heat on the dissociation, denaturation, and aggregation of sesame α-globulin. J Agri Food Chem 27:818–821

Lin LJ, Tzen JTC (2004) Two distinct steroleosins are present in seed oil bodies. Plant Physiol Biochem 42:601–608

Lin LJ, Tai SSK, Peng CC, Tzen JTC (2002) Steroleosin, a sterol-binding dehydrogenase in seed oil bodies. Plant Physiol 128:1200–1211

Lin LJ, Liao PC, Yang HH, Tzen JTC (2005) Determination and analyses of the N-termini of oil-body proteins, steroleosin, caleosin and oleosin. Plant Physiol Biochem 43:700–776

Liu Z, Saarinen NM, Thompson LU (2006) Sesamin is one of the major precursors of mammalian lignans in sesame seed (*Sesamum indicum*) as observed in vitro and in rats. J Nutr 136(4):906–912

Liu TH, Chyan CL, Li FY, Chen YJ, Tzen JTC (2011) Engineering lysine-rich caleosins as carrier proteins torenderbiotin as a hapten onartificial oil bodies for antibody production. Biotechnol Prog 27:1760–1767

Næsted H, Frandsen GI, Jauh GY, Hernandez-Pinzon I, Nielsen HB, Murphy DJ, Rogers JC, Mundy J (2000) Caleosins: Ca^{2+} binding proteins associated with lipid bodies. Plant Mol Biol 44:463–476

Namiki M (1995) The chemistry and physiological functions of sesame. Food Rev Int 11:281–329

Nath R, Giri KV (1957a) Physico-chemical investigations on indigenous seed proteins. I. Studies on the solubilization of nitrogenous constituents of sesame and characterization of its proteins by electrophoresis. J Sci Ind Res 16C:5

Nath R, Giri KV (1957b) Physico-chemical investigations on indigenous seed proteins. II. Fractionation, isolation and eletrophoretic characterization of sesame globulin. J Sci Ind Res 16C:51

Nath R, Rao KH, Giri KV (1957) Physico-chemical investigations on indigenous seed proteins. III. Amino acid composition of sesame seed globulin. J Sci Ind Res 16C:228

Nishimura N, Okubo K, Shibasaki K (1979) Chemical and physical properties of 13S globulin, the major protein in sesame seeds. Cereal Chem 56:239–242

Okubo K, Nishimura N, Shibasaki K (1979a) Composition of sesame seed protein components and purification of the main globulin. Cereal Chem 56:100–104

Okubo K, Nishimura N, Shibasaki K (1979b) Separation of the 13S globulin in sesame seeds into two groups of acidic and basic subunits, and their physiocochemical properties. Cereal Chem 56:317–320

Pasaribu B, Chung TY, Chen CS, Wang SL, Jiang PL, Tzen JTC (2014) Identification of caleosin and two oleosin isoforms in oil bodies of pine megagametophytes. Plant Physiol Biochem 82:142–150

Pasaribu B, Chung TY, Chen CS, Jiang PL, Tzen JTC (2016) Identification of steroleosin in oil bodies of pine megagametophytes. Plant Physiol Biochem 101:173–181

Pasaribu B, Chen CS, Liao YK, Jiang PL, Tzen JTC (2017) Identification of caleosin and oleosin in oil bodies of pine pollen. Plant Physiol Biochem 111:20–29

Peñalvo JL, Heinonen SM, Aura AM, Adlercreutz H (2005) Dietary sesamin is converted to enterolactone in humans. J Nutr 135(5):1056–1062

Peng CC, Tzen JTC (1998) Analyses of the three essential constituents of oil bodies in developing sesame seeds. Plant Cell Physiol 39:35–42

Peng CC, Lin IP, Lin CK, Tzen JTC (2003) Size and stability of reconstituted sesame oil bodies. Biotechnol Prog 19:1623–1626

Peng CC, Chen JCF, Shyu DJH, Chen MJ, Tzen JTC (2004a) A system for purification of recombinant proteins in *Escherichia coli* via artificial oil bodies constituted with their oleosin-fused polypeptides. J Biotechnol 111:51–57

Peng CC, Shyu DJH, Chou WM, Chen MJ, Tzen JTC (2004b) Method for bacterial expression and purification of sesame cystatin via artificial oil bodies. J Agri Food Chem 52:3115–3119

Plietz P, Damaschun G, Zirwer D, Gast K, Schwenke KD, Prakash V (1986) Shape and quaternary structure of alpha-globulin from sesame (*Sesamum indicum* L.)

seed as revealed by small angle x-ray scattering and quasi-elastic light scattering. J Biol Chem 261:12686–12691

Prakash V, Nandi PK (1978) Isolation and characterization of α-globulin of sesame seed (*Sesamum indicum* L.). J Agri Food Chem 26:320–323

Prakash V, Narasinga Rao MS (1990) Hydrogen ion equilibria of α-globulin from *Sesamum indicum* L. J Agri Food Chem 38:612–615

Rajendran S, Prakash V (1988) Isolation and characterization of β-globulin low molecular weight protein fraction from sesame seeds (*Sesamum indicum* L.). J Agri Food Chem 36:269–275

Ritthausen H (1880) Ueber die eiweisskÅrper veschiedener oelsamen. Pfluger's Arch Physiol 21:81–104

Shewry PR (1995) Plant storage protein. Biol Rev 70:375–426

Shih YE, Lin YC, Chung TY, Liu MC, Chen GH, Wu CC, Tzen JTC (2017) In vitro assay to estimate tea astringency via observing flotation of artificial oil bodies sheltered by caleosin fused with histatin 3. J Food Drug Anal 25:828–836

Sinha NK, Sen A (1962) Physico-chemical properties of crystalline globulin of sesame seeds. Trans Bose Res Inst 25:37

Suh MC, Kim MJ, Hur CG, Bae JM, Park YI, Chung CH, Kang CW, Ohlrogge JB (2003) Comparative analysis of expressed sequence tags from *Sesamum indicum* and *Arabidopsis thaliana* developing seeds. Plant Mol Biol 52:1107–1123

Tai SSK, Wu LSH, Chen ECF, Tzen JTC (1999) Molecular cloning of 11S globulin and 2S albumin, the two major seed storage proteins in sesame. J Agri Food Chem 47:4932–4938

Tai SSK, Lee TTT, Tsai CCY, Yiu TJ, Tzen JTC (2001) Expression pattern and deposition of three storage proteins, 11S globulin, 2S albumin, and 7S globulin in maturing sesame seeds. Plant Physiol Biochemistry 39:981–992

Tai SSK, Chen MCM, Peng CC, Tzen JTC (2002) Gene family of oleosin isoforms in sesame seed oil bodies and their structural stabilization to reconstituted oil bodies. Biosci Biotechnol Biochem 66:2146–2153

Tashiro T, Fukuda Y, Osawa T, Namiki M (1990) Oil and minor components of sesame (*Sesamum indicum* L.) strains. J Am Oil Chem Soc 67:508–511

Tzen JTC (2012) Integral proteins in plant oil bodies. ISRN Botany Article ID 263270

Tzen JTC, Lai YK, Chan KL, Huang AHC (1990) Oleosin isoforms of high and low molecular weights are present in the oil bodies of diverse seed species. Plant Physiol 94:1282–1289

Tzen JTC, Cao YZ, Laurent P, Ratnayake C, Huang AHC (1993) Lipids, proteins, and structure of seed oil bodies from diverse species. Plant Physiol 101:267–276

Tzen JTC, Peng CC, Cheng DJ, Chen ECF, Chiu JMH (1997) A new method for seed oil body purification and examination of oil body integrity following germination. J Biochem 121:762–768

Tzen JTC, Chuang RLC, Chen JCF, Wu LSH (1998) Coexistence of both oleosin isoforms on the surface of seed oil bodies and their individual stabilization to the organelles. J Biochem 123:319–324

Tzen JTC, Wang MMC, Chen JCF, Lin LJ, Chen MCM (2003) Seed oil body proteins: oleosin, caleosin, and steroleosin. Curr Topics Biochem Res 5:133–139

Wu LSH, Hong GHH, Hou RF, Tzen JTC (1999) Classification of the single oleosin isoform and characterization of seed oil bodies in gymnosperms. Plant Cell Physiol 40:326–334

Wu WH, Kang YP, Wang NH, Jou HJ, Wang TA (2006) Sesame ingestion affects sex hormones, antioxidant status, and blood lipids in postmenopausal women. J Nutr 136(5):1270–1275

Yuno N, Natoba T, Hirose M, Hasegawa K (1986) Subunit structure of sesame 13S globulin. Agric Biol Chem 50:983–988

Classical Genetics of Sesame

4

Libin Wei, Hongmei Miao, Yinghui Duan, and Haiyang Zhang

Abstract

Improving levels of yield potential, seed quality, resistance to biotic and abiotic stresses, and the adaptation to cultivation mechanization are still the main breeding objectives in sesame in the past decades. To understand the species-specific genetic characteristics of sesame, we review the progress of classical genetics of the key traits including the phenology characters related to the growth and development, yield-related traits, seed quality traits, stress tolerance, and other important morphological and agronomic traits in sesame in this chapter. Especially, plenty of classical genetic studies of the quantitative traits in different genotypes and environments presented by worldwide sesame scientists are reviewed. The results of the classical genetic analysis provide the theoretical basis for elite alleles collection and molecular marker-assisted selection (MAS) in sesame breeding.

4.1 Introduction

Sesame is an annual oilseed crop with long cultivation history. Under various environments, sesame presents high adaptation and great genetic variability, and thus is regarded as an ideal crop for genetic research (Banerjee and Kole 2009a; Zhang et al. 2019). According to the record of Joshi (1961), genetic studies of morphological and agronomic traits in sesame initiated since 1930s. Basic genetic analysis of some key quality traits, such as stem color, branching type, stem fasciation, stem hairiness, leaf arrangement, flower color, capsule carpel number, and seed coat color in sesame was previously studied (Langham 1944, 1945a, b, 1946, 1947). Subsequently, in order to realize high and stable seed yield, some scientists focused on the genetic analysis of the yield related traits in sesame (Ashri 1981). In recent years, genetic correlation and combining ability among the main yield components in sesame are still the hot research points (Liu et al. 1980; Ashri 1998; Gnanasekaran et al. 2008; Banerjee and Kole 2009b; Sumathi and Muralidharah 2010a; 2014). Till now, tens of yield components, such as plant height, number of primary branches, number of capsules per plant, capsule length, number of seeds per capsule, 1000-seed weight, seed dormancy, days to maturity, and other quantitative traits are determined with the various contributions to the final seed yield of sesame plantlet

L. Wei · H. Miao · Y. Duan · H. Zhang (✉)
Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, People's Republic of China
e-mail: zhanghaiyang@zzu.edu.cn

H. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-98098-0_4

(Duc Pham et al. 2010). The findings of the classical genetic analyses of the above traits provide the solid basis for molecular genetic analysis in sesame (Zhang et al. 2016; Mei et al. 2017).

To date, the molecular genetic studies of many agronomic traits in sesame are being systematically performed. However, gene interaction type analyses of different crosses for the quantitative or qualitative traits are still required, in order to elucidate the genetics of characters and to improve the molecular genetics and breeding techniques in sesame (Hossain et al. 2010; Sumathi and Muralidharan Sumathi and Muralidharan 2010). In this section, we summarize the main findings of the classical genetic analyses about key traits including growth and development, yield components, seed quality, disease resistance, abiotic tolerance, and other botanic traits in sesame.

4.2 Genetic Analysis of Morphological Traits

4.2.1 Root Morphology Traits

Root is a vital organ for plants to tolerate water stresses (Passioura 1982; Armenda-Soto et al. 1983; Shashidhar et al. 1990). Satish (2013) investigated seven root morphological traits including root number per plant, maximum root length (cm), root volume (cc), root fresh weight (g), root dry weight (g), root to shoot length ratio (%), and root to shoot dry weight ratio (%) and two shoot morphological traits including shoot fresh weight (g) and shoot dry weight (g) among the six generations of two crosses. The genetic architecture of the seven root morphological traits is also summarized as below. With regard to the root number per plant trait, highly significant and negative additive, dominant, and additive × additive gene effects were detected in both crosses; meanwhile, the additive × dominance and dominance × dominance gene effects presented highly significant and positive in both crosses. Moreover, dominance × dominance interaction was found in high range and the duplicative type of interaction presented. In the cross MT-19-03 × AT-87, additive gene effect was significant and positive. In both crosses, predominance of dominance and dominance × dominance type of epistatic interactions was evident, which suggested the high efficiency of recombination technique for root trait related breeding in sesame.

As regards to the maximum root length trait, the additive, dominant, and additive × additive gene interaction was found significant and negative in both crosses, and the additive × dominant and dominance × dominance type interaction was significant and positive. Meanwhile, the magnitude of dominance × dominance interaction was high, and duplicate interaction existed in both crosses. All the digenic interaction effects were significant in both crosses, indicating the complexity of the root length trait in sesame. Moreover, the results suggested that the dominance and dominance-based interactions would be rewarding for new line selection.

As regards to the root volume trait, significant and negative additive and dominance × dominance gene effects were noticed in both crosses. Dominant, additive × additive, and additive × dominance gene effects were found significant and positive. Meanwhile, dominance followed by additive × additive type of epistasis interaction was higher. Duplicate interaction was also evident in both crosses. The results indicated the predominance of additive, dominance, and epistatic interaction in the inheritance of root volume.

With regard to the root fresh weight trait, additive gene effect was significant and positive in the cross TKG-314 × CST-2001-05, but significant and negative in the cross MT-19-03 × AT-87. Dominant gene effect was highly significant but negative, while the dominance × dominance interaction was highly significant and positive in both crosses. Moreover, the cross MT-19-03 × AT-87 showed the significant negative additive × additive and additive × dominance interactions. The scale of the dominance × dominance interaction was higher, compared to other gene effects. Duplicative

interaction was presented in both crosses. The results showed that all the digenic interactions were significant in the cross MT-19-03 × AT-87 in the negative direction, while the magnitude of dominance × dominance interaction was higher in desirable direction in the cross TKG-314 × CST-2001-05.

Regarding the root dry weight trait, both crosses exhibited highly significant and negative dominance gene effect, and significant and negative additive × additive gene effects. Additive and additive × additive interactions were significant and negative for the cross TKG-314 × CST-2001-05, while both the above gene effects were significant and positive in the cross MT-19-03 × AT-87. Dominance × dominance type of interaction was significant and positive in both the crosses. Duplicate interaction was also detected in these crosses. For the root dry weight trait, the additive, dominance, and epistatic interactions were important in the inheritance of this trait, but dominance and additive × dominance type interaction was of higher magnitude in negative direction. The results suggested the high possibility of screening the genotype with higher root dry weight in the advanced generations.

As to the root to shoot length ratio trait, the cross TKG-314 × CST-2001-05 (−0.07) showed significant but negative additive gene effects, while positive additive gene effect in the cross MT-19-03 × AT-87 (0.04) was not significant. For both the crosses, dominant and additive × additive interaction was significant but negative. Significant and negative additive × dominance gene effects were found in the cross MT-19-03 × AT-87 (−0.11). Dominance × dominance gene interaction was highly significant and positive in both the crosses. Similarly, all the digenic interactions were significant in both the crosses. The results reflected the complexity of the inheritance of root to shoot length ratio. In the extended model, the magnitude of dominance and dominance-based interactions with duplicate type epistasis was high in both the crosses. Thus the genetic basis of dominance and dominance-based interactions suggested that heterosis breeding would be effective for the root to shoot length ratio trait.

Regarding the root to shoot dry weight ratio trait, dominance and additive × additive interaction were significant and negative in both crosses. The additive gene effects of the cross TKG-314 × CST-2001-05 were significant and negative. Meanwhile, the dominance × dominance interaction was significant and positive in both crosses for the trait. Duplicate interaction was also detected in the crosses.

4.2.2 Stem Hairiness

Stem hairiness is an important trait in sesame for its relation with the natural defense against biotic and abiotic stresses. Previous reports indicated that the inheritance pattern of stem hairiness trait was relatively simple. Baydar and Turgut (2000) reported that hairiness was dominant over the hairless trait, which accorded with the segregation ratio of 3:1 in the F_2 generation. Thus the hairiness trait is regulated by a single gene. Falusi (2000) also reported a similar monogenic inheritance in hairiness of stem and petiole in the interspecific cross between *S. indicum*, and *S. prostratum*. Subsequently, Falusi et al. (2002) studied the inheritance of stem and petiole hair density and found that medium hair density trait was dominant over thin long hair trait. Strong hairiness was dominant over light hairiness with ratio of 3:1. Similarly, Yol and Uzun (2011) studied the inheritance of hairiness on stem, leaf, and capsule in sesame. The results proved that the hairiness trait in sesame is controlled by a single dominant gene.

4.2.3 Petiole and Nectary Color

The color of petiole and nectary color of the worldwide sesame germplasm accessions presents diverse results. Thus only few reports focus on the inheritance pattern of petiole and nectary color in sesame. Van Rheenen (1970) studied the inheritance of petiole and nectary color of local

races and pointed out that the mode of inheritance of the traits was monogenic.

4.2.4 Seed Dormancy

Compared with the cultivated sesame, wild *Sesamum* species such as *S. mulayanum* always show deep seed dormancy (Tanesaka et al. 2012). The seeds of *S. indicum* and F_1 of interspecific cross *S. indicum* × *S. mulayanum* showed good germination, while *S. mulayanum* and F_1 of the backcross presented the deep dormant character. Combined with the maternal inheritance of seed coat trait, the seed dormancy of *S. mulayanum* was attributed to the seed coat structure (coat enhanced dormancy). The dormancy depth of the F_3 seeds of *S. indicum* × *S. mulayanum* varied. The ratio of the seeds with deep dormancy (<50% germination) and those with no or shallow dormancy ($\geq$ 50% germination) accorded with the expected ratio of 3:1, which indicated that seed dormancy trait is controlled by a single dominant gene (Tanesaka et al. 2012). Interestingly, the purple pigmentation of the corolla of the interspecific cross was also controlled by a single dominant gene. Investigation of F_2 showed that the purple pigmentation trait presented the linkage relation with seed dormancy.

4.3 Genetic Analysis of the Phenology Characteristics Related to the Growth and Development

For sesame, the life cycle from planting to maturity (for normal harvest) varies from 77 d to 144 d (Langham 2018). According to the description of Langham (2007, 2008, 2018), the cycle of sesame phenology includes four phases, i.e., (1) vegetative stage: from planting to start of flowers; (2) reproductive stage: from start primary flowering to flower termination; (3) ripening stage: from flower termination to physiological maturity; and (4) drying stage: from physiological maturity to direct harvest. To date, except for a few farmers in the US and several Latin American countries where the mechanized harvest technologies are applied, most farmers of the world usually harvest sesame by the end of ripening stage by manual or semi-mechanization. Thus the cycle of sesame here refers to the whole life length from planting to maturity. The traits about 'days to first flowering', 'days to 50% flowering', and 'days to maturity' reflect the important phenology characters in sesame.

4.3.1 Days to First Flowering

The trait of days to first flowering is an important trait and always reflects the growth and development rhythm of a sesame line under a specific environment. Murty (1975) pointed out that the general combining ability variance for 'days to flowering' trait was larger than the specific combining ability variance indicating the predominance of additive gene action. Meanwhile, some reports showed that both additive and non-additive genetic variance were equally important for this trait (Das and Gupta 1999). For instance, Bakheit et al. (2001) reported that the additive × additive type was the most important epistatic effect for 'days to the first flowering' trait. Recently, Sumathi and Muralidharan (2014) investigated the days to first flowering and other nine important traits of the six generations derived from the six crosses (Table 4.1). The results indicated that the days to first flowering of the six varieties varied from 51.14 d to 40.22 d. For days to first flowering, the additive component, dominant component, and their interactions were found in the crosses. In addition, the cross, TMV × KS 99037 presented the additive × additive gene action for the trait, while TMV 4 × KS 99037 was the additive as well as additive × additive interaction effects along with duplicate epistatis. The results reflected the various types of gene effects on days to first flowering trait among different crosses.

Table 4.1 Mean performance of several key agronomic traits of the six generations of different crosses in sesame during 2005–2006 in Coimbatore, India

Cross	Generation					
	P_1	P_2	F_1	F_2	BC_1	BC_2
Days to first flowering						
TMV4 × KS99037	51.14 ± 0.22	48.73 ± 0.47	46.97 ± 0.24	47.03 ± 0.67	49.6 ± 0.34	47.92 ± 0.66
TMV4 × KS990813	51.14 ± 0.22	42.23 ± 0.20	45.31 ± 0.58	48.41 ± 0.23	46.37 ± 0.62	46.59 ± 0.58
TMV4 × KS990153	51.14 ± 0.22	40.22 ± 0.25	45.26 ± 0.70	48.08 ± 0.41	45.73 ± 0.89	46.00 ± 0.84
TMV5 × KS99037	50.09 ± 0.18	48.73 ± 0.47	41.96 ± 0.90	46.82 ± 0.32	43.00 ± 0.44	44.84 ± 0.76
TMV5 × KS990813	50.09 ± 0.18	42.23 ± 0.20	39.95 ± 0.41	41.81 ± 0.34	41.50 ± 0.58	41.15 ± 0.54
TMV5 × KS990153	50.09 ± 0.18	40.22 ± 0.25	41.10 ± 0.54	42.24 ± 0.36	48.45 ± 0.72	43.29 ± 0.93
Days to maturity						
TMV4 × KS99037	99.71 ± 0.29	95.79 ± 0.81	90.83 ± 1.29	93.47 ± 0.48	98.8 ± 0.56	95.46 ± 1.25
TMV4 × KS990813	99.71 ± 0.29	84.19 ± 0.43	91.09 ± 1.05	96.00 ± 0.38	92.83 ± 1.30	93.59 ± 1.19
TMV4 × KS990153	99.71 ± 0.29	80.57 ± 0.35	92.51 ± 1.19	96.22 ± 0.71	91.92 ± 1.52	92.50 ± 1.65
TMV5 × KS99037	100.01 ± 0.23	95.79 ± 0.81	84.57 ± 1.66	93.42 ± 0.61	86.04 ± 0.92	90.40 ± 1.48
TMV5 × KS990813	100.01 ± 0.23	84.19 ± 0.43	80.12 ± 0.66	84.26 ± 0.6	83.27 ± 1.07	83.62 ± 1.06
TMV5 × KS990153	100.01 ± 0.23	80.57 ± 0.35	82.94 ± 0.99	84.74 ± 0.68	96.73 ± 1.36	86.71 ± 1.79
Plant height						
TMV4 × KS99037	94.35 ± 1.57	89.87 ± 2.50	91.08 ± 2.58	90.13 ± 0.90	108.30 ± 2.95	95.19 ± 3.05
TMV4 × KS990813	94.35 ± 1.57	71.41 ± 1.78	103.47 ± 2.91	96.21 ± 1.17	74.83 ± 3.07	91.05 ± 2.34
TMV4 × KS990153	94.35 ± 1.57	62.01 ± 1.98	88.57 ± 2.55	97.52 ± 1.56	106.0 ± 3.88	96.61 ± 4.28
TMV5 × KS99037	106.91 ± 1.85	89.87 ± 2.50	87.43 ± 3.23	86.45 ± 1.55	89.50 ± 2.22	90.66 ± 3.11
TMV5 × KS990813	106.91 ± 1.85	71.41 ± 1.78	96.45 ± 1.72	87.25 ± 1.70	81.68 ± 2.39	69.88 ± 3.15
TMV5 × KS990153	106.91 ± 1.85	62.01 ± 1.98	83.36 ± 2.83	82.78 ± 1.82	94.86 ± 3.10	93.48 ± 4.08
No. of branches						
TMV4 × KS99037	4.38 ± 0.15	2.07 ± 0.17	4.05 ± 0.31	3.99 ± 0.12	4.35 ± 0.41	4.24 ± 0.4
TMV4 × KS990813	4.38 ± 0.15	2.24 ± 0.15	4.64 ± 0.32	4.29 ± 0.16	3.72 ± 0.35	3.73 ± 0.27
TMV4 × KS990153	4.38 ± 0.15	1.81 ± 0.15	5.00 ± 0.28	4.95 ± 0.17	4.51 ± 0.40	4.41 ± 0.49

(continued)

Table 4.1 (continued)

Cross	Generation					
	P_1	P_2	F_1	F_2	BC_1	BC_2
TMV5 × KS99037	5.41 ± 0.20	2.07 ± 0.17	4.27 ± 0.26	3.88 ± 0.14	4.14 ± 0.21	3.49 ± 0.29
TMV5 × KS990813	5.41 ± 0.20	2.24 ± 0.15	4.69 ± 0.24	3.43 ± 0.15	3.59 ± 0.17	2.96 ± 0.25
TMV5 × KS990153	5.41 ± 0.20	1.81 ± 0.15	4.39 ± 0.29	4.12 ± 0.12	4.69 ± 0.38	3.72 ± 0.37
No. capsules per plant						
TMV4 × KS99037	65.29 ± 3.61	48.96 ± 3.44	79.75 ± 5.44	48.86 ± 1.88	104.4 ± 6.94	79.81 ± 5.73
TMV4 × KS990813	65.29 ± 3.61	41.63 ± 2.29	81.47 ± 5.63	68.43 ± 2.48	62.77 ± 5.64	72.27 ± 3.99
TMV4 × KS990153	65.29 ± 3.61	47.12 ± 3.22	65.70 ± 4.33	62.21 ± 2.94	80.35 ± 5.68	65.07 ± 5.72
TMV5 × KS99037	77.27 ± 3.52	48.96 ± 3.44	86.70 ± 6.07	60.66 ± 2.52	64.66 ± 4.42	82.78 ± 6.31
TMV5 × KS990813	77.27 ± 3.52	41.63 ± 2.29	63.02 ± 4.63	64.70 ± 2.51	51.79 ± 3.41	54.50 ± 5.15
TMV5 × KS990153	77.27 ± 3.52	47.12 ± 3.22	57.46 ± 4.31	68.41 ± 3.18	82.59 ± 4.75	88.10 ± 6.54
Capsule length (cm)						
TMV4 × KS99037	2.48 ± 0.03	2.82 ± 0.03	2.58 ± 0.05	2.62 ± 0.02	2.67 ± 0.06	2.65 ± 0.05
TMV4 × KS990813	2.48 ± 0.03	2.59 ± 0.03	2.59 ± 0.04	2.61 ± 0.02	2.60 ± 0.04	2.75 ± 0.05
TMV4 × KS990153	2.48 ± 0.03	2.70 ± 0.03	2.45 ± 0.04	2.51 ± 0.02	2.63 ± 0.06	2.75 ± 0.04
TMV5 × KS99037	2.50 ± 0.03	2.82 ± 0.03	2.69 ± 0.04	2.62 ± 0.02	2.65 ± 0.03	2.59 ± 0.04
TMV5 × KS990813	2.50 ± 0.03	2.59 ± 0.03	2.46 ± 0.04	2.62 ± 0.02	2.58 ± 0.03	2.80 ± 0.03
TMV5 × KS990153	2.50 ± 0.03	2.70 ± 0.03	2.43 ± 0.03	2.42 ± 0.02	2.87 ± 0.05	3.04 ± 0.07
No. of seeds per capsule						
TMV4 × KS99037	44.82 ± 0.82	58.75 ± 0.64	51.19 ± 1.61	50.91 ± 0.63	52.02 ± 2.46	52.76 ± 1.58
TMV4 × KS990813	44.82 ± 0.82	51.37 ± 0.97	50.47 ± 1.57	49.60 ± 0.69	49.63 ± 1.28	56.83 ± 1.96
TMV4 × KS990153	44.82 ± 0.82	51.40 ± 0.87	46.00 ± 1.47	45.07 ± 0.72	49.52 ± 1.65	52.17 ± 1.32
TMV5 × KS99037	43.48 ± 0.98	58.75 ± 0.64	53.16 ± 1.20	48.82 ± 0.58	48.09 ± 1.09	47.90 ± 1.26
TMV5 × KS990813	43.48 ± 0.98	51.37 ± 0.97	44.63 ± 1.19	46.09 ± 0.81	47.38 ± 0.95	53.28 ± 1.40
TMV5 × KS990153	43.48 ± 0.98	51.40 ± 0.87	42.03 ± 1.54	43.34 ± 0.88	54.39 ± 1.12	62.42 ± 0.86
100 seed weight (g)						

(continued)

Table 4.1 (continued)

Cross	Generation					
	P_1	P_2	F_1	F_2	BC_1	BC_2
TMV4 × KS99037	0.33 ± 0.006	0.28 ± 0.007	0.29 ± 0.010	0.29 ± 0.003	0.32 ± 0.010	0.32 ± 0.009
TMV4 × KS990813	0.33 ± 0.006	0.28 ± 0.004	0.30 ± 0.010	0.31 ± 0.004	0.30 ± 0.007	0.29 ± 0.007
TMV4 × KS990153	0.33 ± 0.006	0.27 ± 0.004	0.29 ± 0.007	0.29 ± 0.005	0.29 ± 0.011	0.28 ± 0.008
TMV5 × KS99037	0.33 ± 0.005	0.28 ± 0.007	0.29 ± 0.010	0.29 ± 0.003	0.27 ± 0.005	0.31 ± 0.009
TMV5 × KS990813	0.33 ± 0.005	0.28 ± 0.004	0.30 ± 0.007	0.30 ± 0.003	0.30 ± 0.006	0.29 ± 0.007
TMV5 × KS990153	0.33 ± 0.005	0.27 ± 0.004	0.29 ± 0.007	0.28 ± 0.004	0.31 ± 0.100	0.31 ± 0.100
Seed yield per plant (g)						
TMV4 × KS99037	7.91 ± 0.54	4.89 ± 0.45	9.32 ± 1.06	4.53 ± 0.20	11.31 ± 1.62	8.65 ± 0.78
TMV4 × KS990813	7.91 ± 0.54	3.50 ± 0.22	8.65 ± 0.92	7.28 ± 0.34	5.94 ± 0.68	7.60 ± 0.64
TMV4 × KS990153	7.91 ± 0.54	3.63 ± 0.43	6.67 ± 0.62	5.99 ± 0.39	8.52 ± 1.23	6.34 ± 0.82
TMV5 × KS99037	8.64 ± 0.47	4.89 ± 0.45	8.81 ± 0.94	5.16 ± 0.26	5.56 ± 0.50	9.52 ± 1.11
TMV5 × KS990813	8.64 ± 0.47	3.50 ± 0.22	6.68 ± 0.71	6.23 ± 0.33	4.93 ± 0.42	5.38 ± 0.74
MV5 × KS990153	8.64 ± 0.47	3.63 ± 0.43	6.09 ± 0.61	6.76 ± 0.50	8.97 ± 1.09	11.13 ± 1.49
Oil content (%)						
TMV4 × KS99037	40.72 ± 0.16	39.78 ± 0.13	40.92 ± 0.15	39.78 ± 0.09	40.74 ± 0.26	41.00 ± 0.23
TMV4 × KS990813	40.72 ± 0.16	39.69 ± 0.15	40.19 ± 0.28	40.07 ± 0.12	40.79 ± 0.37	41.50 ± 0.30
TMV4 × KS990153	40.72 ± 0.16	39.23 ± 0.16	40.01 ± 0.20	40.12 ± 0.15	39.72 ± 0.31	40.43 ± 0.25
TMV5 × KS99037	39.92 ± 0.18	39.78 ± 0.13	41.78 ± 0.38	40.11 ± 0.10	40.02 ± 0.21	40.06 ± 0.21
TMV5 × KS990813	39.92 ± 0.18	39.69 ± 0.15	40.27 ± 0.21	40.05 ± 0.14	40.07 ± 0.19	40.04 ± 0.20
TMV5 × KS990153	39.92 ± 0.18	39.23 ± 0.16	39.64 ± 0.19	39.86 ± 0.12	39.82 ± 0.31	39.99 ± 0.29

Note P_1: Parent 1 (Female); P_2: Parent 2 (Male); All values with ± SE. Modified from Sumathi and Muralidharan (2014)

4.3.2 Days to 50% Flowering

Days to 50% flowering indicates the days to flowering (Zhang and Feng 2006; Langham 2018). Similar to the trait 'days to first flowering', the inheritance of days to 50% flowering trait is also a complicated trait. Satish (2013) found that the days to 50% flowering of 10 sesame varieties ranged from 30.35 d (in TKG 314) to 83.00 d (in CST-2001–2005). In the past three decades, numerous reports reflected the genetics background analysis results of days to 50% flowering trait within various crosses (Table 4.2). According to the results of Satish (2013), the additive, dominance, and dominance × dominance gene effects on this trait were shown in the 3 crosses, i.e., MACSS 1 × MT-20-03, TKG-306 × Selec. (N1 × CO1), and RT-127 × ZC, while additive × dominance gene interaction was significant in the crosses MACSS 1 × MT-20-03 and RT-127 × ZC (Satish 2013). For the F_1 derived from the crosses TKG-306 × Selec. (N1 × Co1) and RT-127 × ZC, the mean of the days to 50 per cent flowering exceeded those of the parents and presented the over-dominance character. Moreover, the dominance and dominance × dominance gene effects were significant in all three crosses. Besides the above types of the additive, dominance, and dominance × dominance gene interaction effects, other effects, such as non-additive gene action and the additive × dominance gene action, have been observed in dozens of crosses (Table 4.2). The complicated genetic basis for this trait in sesame reflects the complicated interaction between genotypes and genotypes × environments.

4.3.3 Days to Maturity

The days to maturity trait is also an important trait in sesame, as it reflects the life cycle characterstics of a sesame variety. However, the definition of 'maturity' varies in different sesame production regions. In China, sesame maturity indicates the physiological maturity stage, by which most of the leaves have shed and the lower capsules have become yellowish and almost dehiscent. In the USA, maturity is the date when 50% of the plants reach physiological maturity and the seeds in 75% capsules of the way up the capsule zone appear the seed line and black tip (Langham 2017). In the description of IPGRI (International Plant Genetic Resources Institute, India), maturity was defined as the number of days from planting or first irrigation until 75% of plants reaching physiological maturity. Meanwhile, UPOV (International Union for the Protection of New Varieties of Plants) (2013) explained that maturity was the date when approximately 50% of sesame plants show the capsule dehiscence on the middle third of the main stem. Later, Langham overviewed the above opinions and defined that 'maturity date' was the date when most of the leaves have shed and the lower capsules show a physiological maturity color (Langham 2018).

For sesame, days to maturity is affected by genotype and environment. Of the 763 worldwide sesame germplasm accessions planted in Pingyu, China in 2017, the days to maturity varied from 75 d to 120 d (Unpublished data, Haiyang Zhang). Especially, dozens of exotic varieties, such as some accessions collected from Africa and Latin America, cannot realize the whole life cycle under the natural growth field environment because of their low adaptability for changed day length and latitude. Similar to the traits 'days to first flowering' and 'days to 50% flowering', 'days to maturity' trait is prone to substantial genotype × environment interactions. According to the description of Joshi (Joshi 1961), late maturity dominates over the early maturity by one pair of genes, even though a few somewhat intermediate types occurred. At present, a great amount of reports reflect that the gene interaction effects of days to maturity are diverse in different sesame crosses. Similar to the trait of days to 50% flowering, besides the additive, dominance, and additive and dominance gene actions, other effects including non-additive gene action and the epistatic effect action also have been observed in dozens of crosses (Table 4.3). Till now, no genes related to

Table 4.2 Summary of genetic analyses of days to 50% flowering in sesame crosses

Genetic effect type	Cross structure	References
Additive gene action	9 cultivars and 36 F_1 progeny	Mcharo (1994)
	An 8 × 8 half diallel cross of sesame	Saravanan et al. (2003)
	14 lines	Saxena and Bisen (2017)
Over-dominance	TKG-306 × Selec. (N1 × Co1) RT-127 × ZC	Satish (2013)
Predominance of additive gene action	8 parents, 28 F_1, 28 F_2, and 4 controls	Aye et al. (2018)
Non-additive gene action	Mutant 224 × Mutant 274 Mutant 40 × Mutant 699 Mutant 207 × Mutant 699 Mutant 40 × Mutant 699	Kumar et al. (2012)
	5 lines, 10 testers, and 50 F_1 hybrids	Chaudhari et al. (2016)
	IS-1162 × MT-75 SI-775 × MT-75 NIC-8401 × NIC-16220	Raikwar (2018)
Predominant non-additive gene action	8 lines and 4 varieties as testers	Sankar and Kumar (2003)
	TMV 5 × IS 184 TNAU 2031 × RT 105	Vidhyavathi et al. (2005)
	P1 × P3 P3 × P5 P4 × P5	El-Bramawy and Shaban (2008)
	5 lines, 6 testers, and 30 hybrids	Sumathi and Muralidharan (2008)
	DS-9 × Ahutil DS-9 × Halitil DS-10 × E-8	Yamanura and Nadaf (2009)
	10 lines, 4 testers, and 40 hybrids	Ramesh et al. (2014)
Additive gene, dominance, and dominance × dominance	MACSS 1 × MT-20-03 TKG-306 × Selec. (N1 × CO1) RT-127 × ZC	Satish (2013)
Additive × dominance gene interaction	MACSS 1 × MT-20-03; RT-127 × ZC	Satish (2013)

days to maturity have been cloned in sesame because of the complicated genetic background. Genome-wide associated studies (GWAS) were applied for single nucleotide polymorphism (SNP) loci screening for the days to maturity trait in recent a few years (Wei et al. 2015) (see the following chapters). We thus believe that the genomics and functional genomics research in sesame will conversely stimulate the genetic analysis of the above traits.

4.4 Genetic Analysis of Yield Components Traits

4.4.1 Yield Components Traits

Yield trait is the most important and complicated quantitative trait for crops. Yield trait usually involves many yield-related components traits and is determined by the interaction of yield

Table 4.3 Statistics of genetics analysis results of days to maturity trait in sesame

Genetic effect type	Cross structure	References
Additive gene action	A 6 × 6 full diallel combination	Kamala (1999)
	6parents, 15 F_1	El-Bramawy and Shaban (2008)
	10 mutant lines, 45 F_1 hybrids	Kumar et al. (2012)
	70 genotypes	Monpara and Khairnar (2016)
	14 advanced varietal lines	Saxena and Bisen (2017)
	8 parents, 28 F_1, 28 F_2, and 4 controls	Aye et al. (2018)
Additive and dominance gene action	TMV 4 × KS 99037 TMV 5 × KS 99153	Sumathi and Muralidharan (2014)
Dominance genetic variance	Biparental progeny BIP I (NKD 1110 × Gowri) and BIP II (1799 × Gowri) and F_3 bulk populations CI and C II	Anuradha (2002)
Non-additive gene action	9 lines, 5 varieties, and 45 hybrids	Kumar and Vivekanandan (2009)
	Gowri-173 × JCT-7 Gowri -173 × RT-54	Deepa and Lokesha (2012)
	5 lines, 10testers, and their 50 F_1	Chaudhari et al. (2016)
	AT 355 × Guj.Til-4 RSE 3 × Guj.Til-2 RSE 3 × Guj.Til-3	Mungala (2017)
	ES-230 × DS-10 ES-230 × SI-775 IS-1162 × MT-75	Raikwar (2018)
Predominance of non-additive gene action	TMV 5 × IS 184 TNAU 2031 × RT 105	Vidhyavathi et al. (2005)
	5 lines, 6 testers, and 30 hybrids.	Sumathi and Muralidharan (2008)
	DS-9 × Ahutil DS-9 × Halitil DS-7 × Halitil	Yamanura and Nadaf (2009)
	10 lines, 4 testers, and 40 hybrids	Ramesh et al. (2014)
Pronounced additive, dominance, and non-allelic interaction	NIC7907 × TMV3	Sharmila and Mahalingam (2007)
Dominance × dominance(l) type interaction	Cianno 13/10 × VRI 1 cross	Sharmila and Mahalingam (2007)
Duplicate epistasis	3 lines,5 testers, and their F1	Ramalingam et al. (1990)
	2 crosses	Pathak and Dixit (1988)

components and environmental conditions (Ebata and Electron 2010; Haruna et al. 2012; Gelalcha and Hanchinal 2013; Agrawal et al. 2017). Moreover, most yield components are also regulated by multiple genetic loci. Thus, to reveal the genetics and molecular regulation mechanism of yield-related traits is always the highlight for crop genetics and breeding research (Monpara 2009; Goudappagoudra et al. 2011; Ibrahim and Khidir 2012; Singh et al. 2012; Wu et al. 2014).

For sesame, tens of traits, such as capsule number per plant (CN), seed number per capsule (SN), thousand seed weight (TSW), plant height (PH), height to the first capsule (HFC), branch number (BN), leaf number per plant (LN), capsule number on the main stem zone, seed number per plant (SN), and the above phenology-related (physiological stage) traits are found correlated with the seed yield (Biabani and Pakniyat 2008; Muhamman et al. 2010; Daniya et al. 2013; Jatothu et al. 2013). Tabatabaei et al. (2011) pointed that SN was positively correlated with CN ($r = 0.79$, $P < 0.01$), seed mass of main stem ($r = 0.49$, $P < 0.01$), seed mass of lateral branches ($r = 0.45$, $P < 0.05$), main stem ($r = 0.41$, $P < 0.05$), and flower color ($r = 0.47$, $P < 0.05$), respectively. According to the correlation analysis results, Akbar et al. (2011) pointed out that the five traits including capsule number per plant, days to maturity, plant height, capsule length, and 1000-seed weight had the significant positive contribution to seed yield. Moreover, other four traits, such as primary branches per plant, secondary branches per plant, capsule width, and seeds per capsule were positive but not statistically significant correlated with seed yield per plant. The other characters related to maturity, days to flowering, and days to 50% flowering showed negative correlation with seed yield. Similarly, Goudappagoudra et al. (2011) performed correlation and path coefficient analysis of the 10 quantitative traits within 120 F_4 progeny. Seed yield per plant was detected to be significantly and positively associated with capsule number ($r = 0.7302$), followed by seed number, branch number per plant, plant height, and 1000-seed weight, respectively. Capsule number per plant, seed number per capsule, and 1000-seed weight showed the high and positive direct effect on seed yield. Meanwhile, Ibrahim and Khidir (2012) performed the genotypic correlation and path coefficient analysis, in order to determine the effects of yield components on the final seed yield using 220 F_5 families derived from 10 cross combinations. Highly significant positive genotypic correlation was also detected between seed yield per plant and the seed yield per unit area and the six traits including plant height, capsule number per plant, primary branch number, height to the first capsule, days to 50% flowering, and days to maturity. Path analysis revealed that number of capsules/plant, 1000-seed weight, and number of seeds/capsule had the highest positive direct effects on seed yield/plant.

Furthermore, 40 natural sesame accessions were used to perform the correlation relationship and path analysis of the 14 yield-related traits (Agrawal et al. 2017). Of which the 6 traits, i.e., days to maturity, branch number per plant, capsule number per plant, seed number per capsule, 1000-seed weight, and protein content, showed significant and positive association with seed yield per plant at genotypic and phenotypic levels. Meanwhile, plant height was found to be positively and significantly correlated with the 5 traits (i.e., branch number per plant, capsule number per plant, days to maturity, internode length, and leaf number per plant). Path analysis results showed that the nine traits, i.e., days to maturity, capsule length, branch number per plant, protein content, branch number per plant, capsule number per plant, seed number per capsule, 1000-seed weight, and leaf length, had the positively direct effect with seed yield, while other 4 traits (i.e., plant height, internode length, leaf number per plant, and days to 50% flowering) presented the negative direct effect. Therefore, to increase the yield in sesame, the elite genotypes related with high branch number per plant, capsule number per plant, seed number per

Table 4.4 Statistics of genetics analysis results of seed yield per plant in sesame

Genetic effect type	Cross structure	References
Additive genetic variance	An 8 × 8 half-diallel cross	Das and Gupta (1999)
	biparental progenies BIP I (NKD 1110 × Gowri) and BIP II (1799 × Gowri) and their F_3 bulk populations CI and C II	Anuradha (2002)
	82 genotypes	Kumhar et al. (2008)
	5 lines, 6 testers, and 30 hybrids.	Sumathi and Muralidharan (2009)
	10 lines, 4 testers, and 40 hybrids	Ramesh et al. (2014)
	70 genotypes	Monpara and Khairnar (2016)
	14 lines	Saxena and Bisen (2017)
	8 parents, 28 F_1, 28 F_2, and 4 controls	Aye et al. (2018)
Predominance of additive gene action	10 varieties and F_1 hybrids	Murty (1975)
	30 lines	Banerjee and Kole (2006)
	50 genotypes	Gawali et al. (2007)
Preponderance of dominance gene action	L5 × T3 L2 × T1 L7 × T1 L6 × T1 L3 × T2 L3 × T3 L4 × T2	El-Satar et al. (2016)
	7 parents and their 21 hybrids	Mothilal and Manoharan (2005)
Additive and dominant gene effects	5 variety crosses and their 6 generations	Gaikwad (2004)
	6 crosses	Chavan et al. (1981)
Non-additive gene action	Dhauri Local × DSS-9 DSS-9 × RT-54	Deepa and Lokesha (2012)
	ES-230 × NIC-8401 IS-1162 × NIC-8401 IS-1162 × MT-75	Raikwar (2018)
	Mutant 224 × Mutant 699 Mutant 181 × Mutant 353 Mutant 9 × Mutant 181 Mutant 207 × Mutant 274	Kumar et al. (2012)
	A 6 × 6 full diallel experiment	Kamala (1999)
	an 8 × 8 half diallel cross	Saravanan et al. (2003)
	5 lines, 6 testers, and 30 hybrids.	

(continued)

Table 4.4 (continued)

Genetic effect type	Cross structure	References
		Sumathi and Muralidharan (2008)
Predominance of non-additive gene action	RT 33 × Guj.Til-10 LIMDI 9 × Guj.Til-1 TC 66 × Guj.Til-2	Mungala (2017)
	TMV 5 × N 32 TMV 5 × IS 184 CO 1 × IS 99 TMV 5 × Si 66	Vidhyavathi et al. (2005)
	DS-7 × TSES-2 DS-9 × TSES-3 DS-9 × TSES-4	Yamanura and Nadaf (2009)
	8 lines and 4 testers	Sankar and Kumar (2003)
	9 lines, 5 varieties, and 45 hybrids	Kumar and Vivekanandan (2009)
	P2 × P5 P5 × P6	El-Bramawy and Shaban (2008)
Additive × additive effect	VRI—1 × T6	Kumar and Ganesan (2004)
Dominance × dominance effects	VS 9510 × Co 1 Cianno 13/10 × VRI 1 Si 1115 × 1 TMV 3	Sharmila and Mahalingam (2007)
Additive × additive epistatic effect	90 triple test cross families and parents, F_1 and F_2 of 4 crosses	Bakheit et al. (2001)
Non-additive gene action and the complementary type of epistasis	3 lines, 5 testers, and their F_1	Ramalingam et al. (1990)

capsule, and 1000-seed weight trait should be chosen and applied.

The above results reflect that some traits such as CN, TSW, and SN consistently present the significant and positive correlation with the seed yield per plant (SY). Thus, the above traits are regarded as the principal yield components in sesame (Liu et al. 1980; Gnanasekaran et al. 2008; Banerjee and Kole 2009b; Sumathi and Muralidharah 2010a; Gangadhara et al. 2012; Ibrahim and Khidir 2012).

In recent years, the genetic architecture of physiological stages, yield, and seed quality-related traits have been investigated in sesame. Most of the above quantitative traits present the complicated gene actions involving additive, dominance, additive × additive, additive × dominance, and dominance × dominance types.

4.4.2 Seed Yield Per Plant

Seed yield per plant is the most important yield trait for sesame. So many studies and findings confirmed the complicated genetic architecture with various gene interaction effects within different parent combinations under different ecological conditions (Table 4.4). Kumar and Ganesan (2004) investigated the genetic action of seed yield per plant using 6 generations of the 5

Table 4.5 Statistics of genetics analysis results of capsule number per plant in sesame

Genetic effect type	Cross structure	References
Additive gene action	An 8 × 8 half-diallel cross	Das and Gupta (1999)
	biparental progenies BIP I (NKD 1110 × Gowri) and BIP II (1799 × Gowri) and their F_3 bulk populations CI and C II	Anuradha (2002)
	Biparental progeny of a cross	Anuradha and Lakshmi (2008)
	82 genotypes	Kumhar et al. (2008)
	5 lines, 6 testers, and 30 hybrids.	Sumathi and Muralidharan (2009)
	5 lines, 10 testers, and their 50 F_1	Chaudhari et al. (2016)
	70 sesame genotypes	Monpara and Khairnar (2016)
	14 lines	Saxena and Bisen (2017)
Additive × additive epistatic effect	TTC1, TTC2, and TTC4	Bakheit et al. (2001)
Additive and dominant gene actions	Six crosses of *Sesamum*	Chavan et al. (1981)
	Two crosses of sesame	Pathak and Dixit (1988)
	Five varietal crosses and their six generations (P_1, P_2, F_1, F_2, B_1, and B_2)	aikwad (G2004)
	VRI-1 × T6 TMV 3 × T6	(Kumar and Ganesan 2004)
	Cianno 13/10 × VRI 1 cross.	Sharmila and Mahalingam (2007)
dominance gene action	L5 × T3 L3 × T2 L3 × T3 L4 × T2	El-Satar et al. (2016)
	5 genotypes, 6 cross combinations, and their 6 generations (P_1, P_2, F_1, F_2, BC_1, and BC_2)	Sumathi and Muralidharan (2014)
Non additive gene action	Dhauri Local × DSS-9 DSS-9 × RT-54	Deepa and Lokesha (2012)
	RT 33 × Guj.Til-10 TC 66 × Guj.Til-2 AT 351 × Guj.Til-4 AT 355 × Guj.Til-4 DPI 1484 × Guj.Til-1	Mungala (2017)
	SI-1147 × NIC-8401 SI-775 × NIC-8401 IS-1162 × DS-10	Raikwar (2018)

(continued)

Table 4.5 (continued)

Genetic effect type	Cross structure	References
	Mutant 9 × Mutant 207 Mutant 181 × Mutant 224 Mutant 224 × Mutant 699 Mutant 274 × Mutant 699	Kumar et al. (2012)
Preponderance of non-additive gene action	TMV 5 × IS 184	Vidhyavathi et al. (2005)
	DS-7 × TSES-4 DS-9 × TSES-3 DS-16 × Halitil	Yamanura and Nadaf (2009)
	P1 × P4 P1 × P5 P2 × P4 P2 × P5 P3 × P6	El-Bramawy and Shaban (2008)
	8genotypes as lines and 4 varieties as testers	Sankar and Kumar (2003)
	5e lines, 6 testers, and 30 hybrids.	Sumathi and Muralidharan (2008)
	9 lines, 5 varieties, and 45 hybrids	Kumar and Vivekanandan (2009)
Predominant additive gene action	10 varieties and F_1 hybrids	Murty (1975)
	9 genotypes and 72 hybrids	Rajaravindran et al. (2000)
	30 lines	Banerjee and Kole (2006)
	50 genotypes	Gawali et al. (2007)
Non-additive gene action	3 lines, 5 testers, and their F_1s	Ramalingam et al. (1990)
	9 cultivars and 36 F_1	Mcharo (1994)
	A 6 × 6 full diallel experiment	Kamala (1999)
	An 8 × 8 half diallel cross of sesame	Saravanan et al. (2003)

crosses and found that the dominance effect (h) was significant for the trait. As to the cross VRI-1 × T6, the positive and significant additive × additive effect was observed and reflected the possibility of realizing the transgressive segregation in advanced generations. Moreover, the additive × dominance and dominance × dominance interaction effects were shown in most of the crosses except for Annamalai × T6, which showed the complementary type of epistasis. Meanwhile, Sharmila and Mahalingam (2007) carried out the generation mean analysis of the four sesame crosses. The six-generation analysis of P_1, P_2, F_1, F_2, BC_1, and BC_2 showed that all the gene interaction effects were non-significant for the cross NIC7907 × TMV3, while the remaining 3 crosses showed the dominance × dominance effects. These results indicated that seed yield per plant is predominantly controlled by

dominance × dominance type interaction effects. The dominance (h) and dominance × dominance gene effects showed the presence of duplicate dominant epistasis for the trait. To reveal the genetics basis of seed yield per plant trait, Satish (2013) analyzed the six generations of the three crosses and found that the additive, dominant, and additive × additive types of interactions were significant and negative in all three crosses. Moreover, additive × dominant type of interaction was found positive in cross TKG-306 × Selec. (N1 × CO1) and negative in cross MACSS 1 × MT-20-03. Dominant × dominant type of interaction was highly significant and positive in cross MACSS 1 × MT-20-03 and RT-127 × ZC. Higher magnitude of dominance × dominance type of interaction was found in all cross. Duplicate type of epistasis was recorded in all the three crosses. All the types of gene effects were significant in cross MACSS 1 × MT-20-03 with duplicative epistasis. However, in crosses TKG-306 × Selec. (N1 × Co1) and RT-127 × ZC, both dominance, and additive interactions were significant.

4.4.3 Capsule Number Per Plant

The capsule number per plant (CN) trait is an important quantitative trait in sesame. For most sesame varieties, inflorescence presents indeterminate, and capsule number is easily affected by environmental conditions and the rhythm of plant growth and development. Of the 763 worldwide sesame germplasm accessions cultured in China (Piangyu) (described in the Chap. 2) in 2017, the CN varied from 9.2 to 96.2 with the standard deviation of 16.54 (Unpublished data, Haiyang Zhang). According to the results of numerous reports summarized for CN trait (Table 4.5), gene interaction effects varied within different crosses and environments (Sharmila and Mahalingam 2007; Satish 2013; Sumathi and Muralidharan 2014). Sumathi and Muralidharan (2014) investigated the CN variation of the generations of the six crosses. Of the six varieties, TMV5 had the highest average CN (77.27), while KS990813 exhibited the lowest number (41.63). In TMV4 × KS99037, TMV5 × ks99153, TMV4 × KS99153, and TMV5 × KS99037, the dominance component was significant and positive for CN trait, similar to the results of Vidhyavathi (2002). In TMV4 × ls99153, additive dominance model was observed. However, both additive and dominance gene actions were found in the cross TMV 4 × KS99037. Meanwhile, Satish (2013) also found that the negative additive gene effects, positive dominance effects, additive × additive gene interaction, and positive additive × dominance interaction, dominant × dominant type interaction were presented in the different crosses. Moreover, higher dominance and dominance × dominance gene interaction presented in all the three crosses, and duplicate gene action was observed in all the crosses.

As shown in Table 4.5, besides the additive, dominance, and both additive and dominance gene interaction effects, other effects such as non-additive gene action and epistatic effects are observed in dozens of crosses. The complicated genetic basis for CN trait in sesame exhibits the complexity of genotype to genotype and the interaction between genotypes and environments.

4.4.4 Thousand Seed Weight

Sesame belongs to small seed-sized crop type. Increasing the seed size or seed weight can contribute to high seed yield per plant and the final productivity. For sesame, the thousand seed weight usually varies from 0.79 to 4.47 g (data not shown, Haiyang Zhang). Previous studies proved that thousand seed weight trait is controlled by substantial genotype × environment interactions in sesame. Some reports indicated that the genetic basis for the TSW trait varied from the additive, dominance, both additive and dominance gene interaction to non-additive gene action, and additive × dominance and dominance × dominance interaction epistatic effect (Table 4.6). The complicated genetic bases for TSW trait in sesame reflect the complex interaction between genotype to genotype and

Table 4.6 Statistics of genetics analysis results of thousand seed weight in sesame

Genetic effect type	Cross structure	References
Additive gene action	9 cultivars and 36 F_1 progeny	Mcharo (1994)
	An 8 × 8 half diallel cross of sesame	Saravanan et al. (2003)
	10 lines, 4 testers, and 40 hybrids	Ramesh et al. (2014)
	5lines, 10 testers, and 50 F_1	Chaudhari et al. (2016)
	14 advanced lines	Saxena and Bisen (2017)
Additive × additive epistatic effect	90 triple test cross families and their parents, F_1 and F_2 in 4 sesame crosses	Bakheit et al. (2001)
Additive and complementary effect	Si 1115/1 × TMV 3	Sharmila and Mahalingam (2007)
Additive, dominance × dominance, and additive × dominance effects and duplicate epistasis	VS 9510 × Co1; NIC 7907 × TMV 3, and Cianno 13/10 × VRI 1	Sharmila and Mahalingam (2007)
Predominant additive gene action	9 genotypes and 72 hybrids	Rajaravindran et al. (2000)
	50 genotypes	Gawali et al. (2007)
Non-additive gene action	An 8 × 8 half-diallel cross	Das and Gupta (1999)
	A 6 × 6 full diallel	Kamala (1999)
	7 parents and 21 hybrids	Mothilal and Manoharan (2005)
	9 lines, 5 varieties, and 45 hybrids	Kumar and Vivekanandan (2009)
	Mutant 9 × Mutant 699 Mutant 51 × Mutant 181 Mutant 224 × Mutant 450 Mutant 353 × Mutant 450	Kumar et al. (2012)
	DSS-9 × JCT-7 RT-54 × MT-75	Deepa and Lokesha (2012)
	TC 66 × Guj.Til-4 RT 33 × Guj.Til-10 TC 66 × Guj.Til-10	Mungala (2017)
Predominant non-additive gene action	8 genotypes as lines and 4 varieties as testers	Sankar and Kumar (2003)
	P4 × P5 P5 × P6	El-Bramawy and Shaban (2008)
	5 lines, 6 testers, and 30 hybrids	

(continued)

Table 4.6 (continued)

Genetic effect type	Cross structure	References
		Sumathi and Muralidharan (2008)
	19 parents and 90 crosses	Yamanura and Nadaf (2009)
Predominant dominance gene action	L7 × T1 L6 × T1 L3 × T2 L3 × T3 L4 × T2	El-Satar (et al. 2016)
Dominance gene action	Biparental progenies BIP I (NKD 1110 × Gowri) and BIP II (1799 × Gowri) and F_3 bulk populations CI and C II	nuradha (A2002)
	Biparental progeny of a cross	Anuradha and Lakshmi (2008)
Additive × dominance and dominance × dominance interaction effects	5 crosses and their 6 generations (P_1, P_2, F_1, F_2, BC_1, and BC_2)	Kumar and Ganesan (2004)

between genotype and environment. Sharmila and Mahalingam (2007) reported that the TWS trait was additive and complementary for the cross 'Si 1115/1' × 'TMV 3', while additive, dominance × dominance, and additive × dominance effects and duplicate epistasis were predominant in the other three crosses. Various gene interactions indicate both fixable and non-fixable gene effects. Genetic architecture of TSW suggests that the transgressive segregation in advanced generations should be possible. Satish (2013) found the significant and negative additive gene effects in all three crosses. In cross RT-127 × ZC, the dominance gene interaction present significant and positive; however, the dominance gene interaction was negative and significant in cross TKG-306 × (N1 × CO1). Meanwhile, the significant and positive additive × additive type was found in crosses MACSS 1 × MT-20-03 and RT-127 × ZC. The cross RT-127 × ZC exhibited significant and negative additive × dominance and dominance × dominance type of gene interaction. Moreover, the mean value of the TSW of F_1 hybrid was intermediate, compared to parents in the crosses MACSS 1 × MT-20-03 and TKG-306 × Selec. (N1 × Co1), indicating the incomplete dominance. In the cross, RT-127 × ZC, the mean value of F_1 was higher than those of parents and presented the over-dominance character. Meanwhile, duplicate type of epistasis for TSW trait was noticed in all three crosses.

4.4.5 Seed Number Per Capsule

Seed number per capsule is another important yield component trait. So many sesame scientists confirmed the complicated inheritance of seed number per capsule trait (Table 4.7). Simple models seem inadequate to explain the genetic base of the trait. Previous reports showed the various gene interaction effects within the different crosses under diverse environments (Kumar and Ganesan 2004; Mungala 2017;

Table 4.7 Statistics of genetics analysis results of seed number per capsule in sesame

Genetic effect type	Cross structure and heredity value?	References
Non-additive gene action	9 cultivars and 36 F_1 progeny	Mcharo (1994)
	A 6 × 6 full diallel	Kamala (1999)
	An 8 × 8 half diallel cross	Saravanan et al. (2003)
	5 lines, 6 testers, and 30 hybrids.	Sumathi and Muralidharan (2008)
	9 lines, 5 varieties, and 45 hybrids	Kumar and Vivekanandan (2009)
	Gowri-173 × Dhuri Local JCT-7 × RT-54	Deepa and Lokesha (2012)
	Mutant 181 × Mutant 207 Mutant 181 × Mutant 224 Mutant 9 × Mutant 207 Mutant 207 × Mutant 224	Kumar et al. (2012)
	DPI 1484 × Guj.Til-10 AT 355 × Guj.Til-2 TC 66 × Guj.Til-1	Mungala (2017)
	ES-230 × IS-1162 SI-775 × NIC-16220 SI-775 × DS-10	Raikwar (2018)
Additive gene action	9 genotypes and 72 hybrids	Rajaravindran et al. (2000)
	30 lines	Banerjee and Kole (2006)
	70 genotypes	Monpara and Khairnar (2016)
	14 advanced lines	Saxena and Bisen (2017)
Additive × additive interaction effect	AnnamaJai × T6 TMV 3 × T6	Kumar and Ganesan (2004)
Preponderance of additive gene action	50 genotypes	Gawali et al. (2007)
Dominance effect	5 crosses and 6 generations (P_1, P_2, F_1, F_2, BC_1, and BC_2)	Kumar and Ganesan (2004)
	Biparental progeny of a cross	Anuradha and Lakshmi (2008)
Additive and dominant gene effects	5 crosses and 6 generations (P_1, P_2, F_1, F_2, BC_1, and BC_2)	Gaikwad (2004)

Saxena and Bisen 2017). Satish (2013) found that additive gene effect was significant and negative in the crosses TKG-306 × Selec. (N1 CO1) and RT-127 × ZC. However, additive × dominance type of interaction was found significant and negative in cross MACSS 1 × MT-20-03 and TKG-306 × Selec. (N1 × CO1). Dominance × dominance interaction presented negative and significant in cross MACSS 1 × MT-20-03. Moreover, two crosses MACSS 1 × MT-20-03 and RT-127 × ZC exhibited duplicate type, while cross TKG-306 × Selec. (N1 × CO1) manifested the complementary type of epistasis for this trait. In the F_1 hybrid, high seed number per capsule was observed in all the three crosses and suggested the overdominance characteristics. Meanwhile, great variation of seed number per capsule in F_2 progeny indicated the high possibility of transgressive segregation. The mean value of BC_1 and

Table 4.8 Statistics of genetics analysis results of plant height in sesame

Genetic effect type	Cross structure	References
Additive gene action	A 6 × 6 full diallel	Kamala (1999)
	9 genotypes and 72 hybrids	Rajaravindran (et al. 2000)
	Biparental progenies BIP I (NKD 1110 × Gowri) and BIP II (1799 × Gowri) and F_3 bulk populations CI and C II	Anuradha (2002)
	An 8 × 8 half diallel cross	Saravanan et al. (2003)
	7 parents and 21 hybrids	Mothilal and Manoharan (2005)
	Biparental progeny of a cross	Anuradha and Lakshmi (2008)
	5 lines, 6 testers, and 30 hybrids	Sumathi and Muralidharan (2008)
	5 lines, 6 testers, and 30 hybrids	Sumathi and Muralidharan (2009)
	14 advanced lines	Saxena and Bisen(2017)
Predominant additive gene action	10 varieties and F_1 hybrids	Murty (1975)
	8 genotypes as lines and 4 varieties as testers	Sankar and Kumar (2003)
	30 lines	Banerjee and Kole (2006)
	IC 96128 × Guj.Til-3 RSE 3 × Guj.Til-3	Mungala (2017)
Dominance effect	5 crosses and 6 generations (P_1, P_2, F_1, F_2, BC_1, and BC_2)	Kumar and Ganesan (2004)
Predominant dominance gene action	L2 × T1 L7 × T1 L6 × T1 L3 × T2 L3 × T3 L4 × T2	El-Satar et al. (2016)
Additive and dominance gene actions	TMV 4 × KS 99037	Sumathi and Muralidharan (2014)
Non-additive gene action	MT-75 × RT-54 JCT-7 × CO-1	Deepa and Lokesha (2012)
	10 lines, 4 testers, and 40 hybrids	Ramesh et al. (2014)
	9 lines, 5 varieties, and 45 hybrids	Kumar and Vivekanandan (2009)
	Mutant 9 × Mutant 353 Mutant 9 × Mutant 224	Kumar et al. (2012)

(continued)

Table 4.8 (continued)

Genetic effect type	Cross structure	References
	Mutant 274 × Mutant 699 Mutant 207 × Mutant 274	
	ES-230 × SI-1147 IS-1162 × NIC-8401 DS-10 × MT-75	Raikwar (2018)
Predominant non-additive gene action	P2 × P5; P5 × P6	El-Bramawy and Shaban (2008)
	TNAU 2031 × IS 99	Vidhyavathi et al. (2005)
	DS-16 × TSES-4 DS-16 × TSES-1 DS-9 × Western	Yamanura and Nadaf (2009)
Dominance × dominance type	MACSS 1 × MT-20-03 TKG-306 × Selec. (N1 × CO1)	Satish (2013)
Dominance and additive × additive	RT-127 × ZC MACSS 1 × MT-20-03 TKG-306 × Selec. (N1 × CO1)	Satish (2013)
Additive × dominance and dominance × dominance effects	VS 9510 × Co 1 cross	Sharmila and Mahalingam (2007)
Additive × dominance effect	Cianno 13/10 × VRI 1 cross	Sharmila and Mahalingam (2007)
Additive × additive epistatic effect	90 triple test cross families and parents, F_1, and F_2 in 4 crosses	Bakheit et al. (2001)
Complementary type of epistasis	3 lines, 5 testers, and F_1	Ramalingam et al. (1990)
Duplicate-type epistasis	All cases	Sharmila and Mahalingam (2007)

BC_2 generation was similar to those of the parents indicating the possibility of maternal effect for the trait exhibition. In the cross, MACSS 1 × MT-20-03, additive, dominance, and epistatic gene interactions were significant, which indicated the ambiguity improvement of the trait. However, the dominance-based interactions were of higher magnitude in a desirable direction.

4.4.6 Plant Height

Plant height is an important trait for sesame, as it is always correlated with the capsule stem length and capsule number. So many results indicated that inheritance of plant height is complicated (Table 4.8). The gene interaction effects are decided on crosses (or genotypes) and the environments (Vidhyavathi et al. 2005; El-Satar et al. 2016; Mungala 2017; Raikwar 2018). Sharmila and Mahalingam (2007) pointed out that the additive × dominance and dominance × dominance effects were significant in the VS 9510 × Co 1 cross, while the additive × dominance effect was significant only in the Cianno 13/10 × VRI 1 cross. In all cases, duplicate-type epistasis was also observed for plant height traits. Moreover, Satish (2013) showed that additive, dominance, and epistatic interactions were highly significant in the inheritance of plant height across the three

Table 4.9 Genetic analysis statistics of other 17 yield-related traits in sesame

Trait	Genetic basis	Cross type	References
1. Capsules per leaf axil	One capsule per leaf axil trait is dominant to three capsules per leaf axil, controlled by single gene	6 genotypes and F_2	Baydar and Turgut (2000)
		4 genotypes and F_2	Yol and Uzun (2011)
2. Locule number	4 locule trait is dominant to multi-locule, controlled by single gene	6 genotypes and F_2	Baydar and Turgut (2000)
		5 genotypes and10 F_2	Padmasundari et al. (2010)
3. Leaf shape	Lobed leaf shape was dominant to simple or ovate leaf shape, controlled by single gene	6 genotypes and F_2	Baydar and Turgut (2000)
4. Capsule length	Additive gene action	Biparental progenies BIP I (NKD 1110 × Gowri) and BIP II (1799 × Gowri) and F_3 bulk populations CI and C II	Anuradha (2002)
		14 advanced lines	Saxena and Bisen (2017)
		8 parents, 28 F_1, 28 F_2, and 4 controls	Aye et al. (2018)
	long capsule trait is dominant to short capsule trait	2 parents and F_1, F_2, and F_3 progeny	Yol et al. (2017)
	Non-additive gene action	9 cultivars and 36 F_1 progeny	Mcharo (1994)
		9 lines, 5 varieties, and 45 hybrids	Kumar and Vivekanandan (2009)
		Mutant 9 × Mutant 40 Mutant 40 × Mutant 224 Mutant 40 × Mutant 181 Mutant 224 × Mutant 274	Kumar et al. (2012)
		RT 33 × Guj.Til-3 DPI 1484 × Guj.Til-1 DPI 1484 × Guj.Til-10	Mungala (2017)
		ES-230 × NIC-16220 SI-775 × NIC-8401	Raikwar(2018)
	Predominant non-additive gene action	5 lines, 6 testers, and 30 hybrids	Sumathi and Muralidharan (2008)
		DS-10 × Western DS-13 × Halitil DS-13 × Ahutil	Yamanura and Nadaf (2009)
		DS-10 × TSES-3 DS-10 × TSES-4 DS-10 × Ahutil	Yamanura and Nadaf (2009)
	Additive × additive epistatic effect	90 triple test cross families and their parents, F_1, and F_2 of 4 crosses	Bakheit et al. (2001)
	Duplicate type of epistasis	3 lines, 5 testers, and F_1	(Ramalingam et al. 1990)

(continued)

Table 4.9 (continued)

Trait	Genetic basis	Cross type	References
	Additive, dominance × dominance, and additive × dominant epistasis type	MACSS 1 × MT-20-03	Satish (2013)
	Additive × additive interaction effect	TMV 5 × T6	Kumar and Ganesan (2004)
	Additive and dominant and additive × additive gene effects	RT-127 × ZC	Satish (2013)
	Additive and additive × dominance interaction	TKG-306 × Selec. (N1 × CO1)	Satish (2013)
	Dominance × dominance type	MACSS 1 × MT-20-03 TKG-306 × Selec. (N1 × CO1)	Satish (2013)
5. Branch number per plant	Branching type is dominant to uniculm trait with segregation ratio of 3:1	TMV 4 × KS 99037 TMV 4 × KS 990813 TMV 5 × KS 99037 TMV 5 × KS990813 TMV 5 × KS99153	Sumathi and Muralidharan (2014)
		6 genotypes and F_2 generations	Baydar and Turgut (2000)
	Additive gene action	14 advanced lines	Saxena and Bisen (2017)
		5 lines, 6testers, and 30 hybrids	Sumathi and Muralidharan (2009)
		82 genotypes	Kumhar et al. (2008)
		7parents and 21 hybrids	Mothilal and Manoharan (2005)
		5 lines, 6 testers, and 30 hybrids.	Sumathi and Muralidharan (2008)
		An 8 × 8 half diallel cross	Saravanan et al. (2003)
		A 6 × 6 full diallel	Kamala (1999)
		An 8 × 8 half-diallel cross	Das and Gupta (1999)
	Predominance of additive gene action	10 varieties and F_1 hybrids	Murty (1975)
		8 genotypes as lines and 4 varieties as testers	Sankar and Kumar(2003)
		30 lines	Banerjee and Kole (2006)
		8 parents, 28 F_1, 28 F_2, and 4 controls	Aye et al. (2018)
	Dominance genetic variance		

(continued)

Table 4.9 (continued)

Trait	Genetic basis	Cross type	References
		Biparental progenies BIP I (NKD 1110 × Gowri) and BIP II (1799 × Gowri) and F_3 bulk populations CI and C II	Anuradha (2002)
		Biparental progeny of a cross	Anuradha and Lakshmi (2008)
	Predominant dominance effect	5 crosses and 6 generations (P_1, P_2, F_1, F_2, B_1, and B_2)	Kumar and Ganesan (2004)
	Predominant non-additive gene action	P2 × P3 P5 × P6	El-Bramawy and Shaban (2008)
		Nine lines, five released varieties, and 45 hybrids	Kumar and Vivekanandan (2009)
	Non additive gene action	DSS-9 × Dhauri Local MT -75 × RT-54	Deepa and Lokesha (2012)
		TC 66 × Guj.Til-2 DPI 1484 × Guj.Til-1 RSE 3 × Guj.Til-10	Mungala (2017)
	Additive x additive epistatic effect	90 triple test cross families and parents, F_1, and F_2 of 4 crosses	Bakheit et al. (2001)
	Dominance x dominance effect	VS9510, × Co 1 NIC 7907 × TMV 3	Sharmila and Mahalingam (2007)
	Both additive × dominance and dominance × dominance effects	Cianno 13/10 × VRI1	Sharmila and Mahalingam (2007)
	Duplicate-type epistasis was complementary	Si 1115/1 × TMV 3	Sharmila and Mahalingam (2007)
	Additive and dominance type, and additive × additive gene effects	MACSS 1 × MT-20-03 RT-127 × ZC	Satish (2013)
		TKG-306 × Selec. (N1 × CO1) RT-127 × ZC	Satish(2013)
	Additive and dominance type and duplicate epistasis type	MACSS 1 × MT-20-03 TKG-306 × Selec. (N1 × CO1) RT-127 × ZC	Satish (2013)
6. Secondary branch number per plant	Additive gene	14 advanced lines	Saxena and Bisen (2017)
		8 parents, 28 F_1, 28 F_2, and 4 controls	Aye et al. (2018)
	Predominance of additive gene action	10 varieties and F_1 hybrids	Murty (1975)
		9 genotypes and 72 hybrids	Rajaravindran et al. (2000)

(continued)

Table 4.9 (continued)

Trait	Genetic basis	Cross type	References
	Non-additive gene effects are more conspicuous	9 cultivars and 36 F_1 progeny	Mcharo (1994)
	Additive gene effects	9 cultivars and 36 F1 progeny	Mcharo (1994)
	Additive × additive epistatic effect	90 triple test cross families and their parents, F_1 and F_2 in 4 sesame crosses	Bakheit et al. (2001)
	Preponderance of dominance gene action	L5 × T3 L2 × T1 L7 × T1 L6 × T1 L3 × T2 L3 × T3 L4 × T2	El-Satar et al. (2016)
	Non-additive genetic component	RT 33 × Guj.Til-10 TC 66 × Guj.Til-10 AT 351 × Guj.Til-4	Mungala (2017)
8. Seed number per plant	Preponderance of additive gene action	A6 parent diallel including reciprocals	Krishna et al. (2002)
9. Number of filled seeds per plant	Preponderance of additive gene action	A 6 parent diallel including reciprocals	Krishna et al. (2002)
10. Number of ill-filled seeds per plant	Non-additive gene action	A6 parent diallel including reciprocals	Krishna et al. (2002)
11. Flower number per plant	Preponderance of additive gene action	A6 parent diallel including reciprocals	Krishna et al. (2002)
12. Reproductive efficiency	Preponderance of additive gene action	A6 parent diallel including reciprocals	Krishna et al. (2002)
13. Photosynthetic rate	Additive genetic variance	An 8 × 8 half diallel cross	Saravanan et al. (2003)
14. Leaf area index	Non-additive variance	An 8 × 8 half diallel cross	Saravanan et al. (2003)
15. Harvest index	Additive genetic variance	An 8 × 8 half diallel cross	Saravanan et al. (2003)
		14 advanced lines	Saxena and Bisen (2017)
	Dominance variance	Biparental progeny of a cross	Anuradha and Lakshmi (2008)
16. Biological yield	Dominance variance.	Biparental progeny of a cross	Anuradha and Lakshmi (2008)
17. Capsule bearing height	Additive gene action	82 genotypes	Kumhar et al. (2008)

crosses. Dominance × dominance interaction type in crosses MACSS 1 × MT-20-03 and TKG-306 × Selec (N1 × CO1) and dominance and additive × additive in RT-127 × ZC, MACSS 1 × MT-20-03, and TKG-306 × Selec. (N1 × CO1) cross was positive and high, whereas the additive, dominance, and additive × dominance interactions were significant in negative direction. Duplicate epistasis type was noticed in all three crosses. In the above three crosses, F_1 presented the overdominant character for their higher values of plant height than those of parents. Interestingly, additive and additive × additive gene effects were significant but opposite in direction with each other. Therefore, selecting dwarf varieties with low height should be effective in all the crosses. However, higher magnitude of dominance × dominance gene effects in crosses MACSS 1 × MT-20-03 and TKG-306 × Selec. (N1 × Co1) and the dominance gene effects in cross RT-127 × ZC indicated the importance of dominance gene effects on plant height.

4.5 Other Yield-Related Traits

Besides the above yield components traits, there are still other traits contributing to the seed yield in sesame. Here we listed the genetic analysis of other 17 traits to supply the genetic mechanism basis for further molecular regulation analysis of the yield trait in sesame (Table 4.9). The inheritance of the above traits such as capsules per leaf axil, locule number, and leaf shape was relatively simple and proved to be controlled by single gene (Nohara 1933; Baydar and Turgut 2000; Padmasundari et al. 2010; Yol and Uzun 2011). Some traits such as capsule length, internode length, height to first capsule, and harvest index are complex with various gene interaction types involving in additive, dominance, and epistatic effects in different combinations (Mcharo 1994; Bakheit et al. 2001; Yamanura and Nadaf 2009; Satish 2013; Mungala 2017; Saxena and Bisen 2017; Aye et al. 2018).

Notably, even though the branch number per plant trait is controlled by complicated gene interaction types including additive, dominance, and epistatic effects in different various cross combinations (Sharmila and Mahalingam 2007; Kumhar et al. 2008; Aye et al. 2018), branching type is monogenic dominant to uniculm type (Baydar and Turgut 2000; Sumathi and Muralidharan 2014). Furthermore, many scientists pointed out that the additive gene action was predominant for the seven traits including secondary branch number per plant, total seed number per plant, filled seed number per plant, flower number per plant, photosynthetic rate, capsule stem length, and reproductive efficiency (Murty 1975; Rajaravindran et al. 2000; Saravanan et al. 2003; Kumhar et al. 2008; Saxena and Bisen 2017; Aye et al. 2018) In addition, non-additive gene effects were more conspicuous for the height to first branch and ill-filled seed number per plant trait (Mcharo 1994).

4.6 Seed Quality Traits

Sesame seeds contain abundant unsaturated fatty acids, proteins, and antioxidants such as lignan and vitamin E. The analysis of genetic basis of oil content, protein content, lignan content, and

Table 4.10 Resistance scale of sesame against *Fusarium* wilt

Grade	Disease index (DI)	Resistance level
1	$0 \leq DI \leq 15$	Highly resistant (HR)
3	$15 < DI \leq 30$	Resistant (R)
5	$30 < DI \leq 55$	Moderately resistant (MR)
7	$55 < DI \leq 70$	Susceptible (S)
9	$70 < DI \leq 100$	Highly susceptible (HS)

Modified from Qiu et al. (2014)

Table 4.11 Grade scale of powdery mildew disease in sesame

Disease grade	Description
0	No lesions or specks
1	Small-sized powdery specks infecting less than 1% leaf area
3	Enlarged irregular powdery growth covering 1–5% leaf area
5	Powdery growth to form big patches covering 5–25% leaf area
7	Powdery growth covering 25–50% leaf area followed by yellowing
9	100% leaf area covered with powdery growth and infected leaves yellowing and dropping

Modified from Rao et al. (2012))

their components' ratio is a hot topic in sesame genetic research. In the past few decades, in order to clarify the genetic behavior of the high quality of sesame oil and protein, sesame researchers focused on the genetic basis and the inheritance of the oil content and protein content and their correlation, as well as digging out related genes and molecular markers (Jin et al. 2001; Chun et al. 2003; Leduc et al. 2006; Magni et al. 2010; Wei et al. 2013; Li et al. 2014).

4.6.1 Oil and Protein Content

Oil content (OC) and protein content (PO) are complex quantitative traits and are affected by genotype and environment. Yermanos et al. (1972) assessed 721 worldwide sesame varieties, and found that the oil and protein contents ranged from 40.4% to 59.8% and from 19% to 31%, respectively. Li et al. (2014) evaluated 369 worldwide sesame germplasm accessions under five environments. The oil content varied from 27.89% to 58.73% with an average of 51.34%, while the protein content varied from 16.72% to 27.79% with an average of 21.19%. The results proved that the oil and protein contents of the natural germplasm population varied obviously. Meanwhile, the OC and PC traits in sesame were significantly influenced by genotype and genotype × environment, as the V(G)/V(P) reached up to 76.2 and 50, respectively (Li et al. 2014). Moreover, The OC is easily affected by genotype, as the broad-sense heritability (H^2) values of OC trait are higher than those of PC trait. The high correlation coefficient ($r = -0.66$) between OC and PC indicates that the OC and PC traits have the significant and negative correlation (Li et al. 2014).

In order to analyze the inheritance of oil and protein contents, Culp (1959) primarily investigated the phenotype of the cross N124 × K8. The oil and protein content presented the high broad-sense heritability value of about 50% and 60%, respectively. One or two genes were estimated to determine the inheritance of oil content, while one gene determined the protein content. Murty (1975) examined the oil and protein contents and other six agronomic traits within a complete diallel cross of 10 sesame varieties. The oil and protein contents of the 10 parents varied from 40.53−48.11%, to 16.83−24.48%, respectively. Combining ability analyses revealed that the specific combining ability was high for protein content, while the heterosis was insignificant for oil content. Banerjee and Kole (2009a) investigated the genetic architecture of the five physiological parameters and oil content and oil yield per plant in seven sesame genotypes and their 21 F_{1s} and F_{2s} progeny. The results showed that the additive gene action for OC was predominant in F_1, while non-additive gene action in F_2. Fatteh et al. (1982) studied the genetic variation of the oil content using a diallel cross set among the 6 sesame lines. The non-additive type of genetic variation appeared contributing to oil content. Furthermore, Chandramony and Nayer (1988) performed diallel analysis of six varieties and indicated that oil content was controlled by non-additive gene action. Mansouri and Ahmadi (1998) studied the combining ability for seven sesame lines crossed in diallel style and

found that the genetic variance for oil content resulted from the additive gene action.

Das and Gupta (1999) reported that both additive and non-additive genetic variance were equally important for oil content. Saravanan et al. (2000) reported that both SCA (specific combining ability) and GCA (general combining ability) variance were significant for oil content. Sankar and Kumar (2003) observed the significant additive and non-additive genetic effects for oil content. However, some scientists reported that additive genetic variance was more important for oil content (Saravanan et al. 2003; Vidhyavathi et al. 2005). Babu et al. (2004) estimated the genetic characters of oil content in 18 white-seeded sesame genotypes. Low heritability and narrow variability range in the advanced generations were observed and indicated the preponderance of non-additive gene action. Thereby the potential to improve the oil content trait through advanced generation selection would be limited. Kumar and Vivekanandan (2009) and Kumar et al. (2012) also reported the predominance of non-additive gene action in the inheritance of oil content. Meanwhile, El-Bramawy and Shaban (2008) studied the gene action for the oil content in 6 × 6 half-diallel progenies (F_1) and detected the preponderance of non-additive genetic variance. Overdominance and the positive and negative gene effects were symmetrical or nearly symmetrical with respect to this trait. Sumathi and Muralidharan (2009) analyzed the gene action and the heterosis using 11 genotypes including five lines and six testers. SCA variance was greater than GCA variance for oil content trait, suggesting the predominance of non-additive components. According to line × tester analysis of 19 parents and 90 crosses in sesame, Yamanura and Nadaf (2009) proved that the oil content trait was determined by the predominance of non-additive gene action and could be improved by biparental mating. Sumathi and Muralidharan (2014) performed a field experiment with five sesame genotypes and six cross combinations and the six generations to estimate the gene effects for oil content. None of the crosses exhibited additive genetic variance for the trait. Meanwhile, EL-Satar et al. (2016) analyzed the gene action for the oil content with 21 F_1 crosses of seven lines and three testers. The specific combining ability variance was much higher than the general combining ability variance, indicating the preponderance of dominance gene action for the oil content trait.

4.6.2 Sesamin and Sesamolin Content

Sesamin and sesamolin are the two major components of oil-soluble lignans in sesame seeds (Wang et al. 2012; Zhang et al. 2019). Ogata and Kato (2016) reported that the sesamin and sesamolin contents in sesame fit an additive-dominant model without the influence of epistasis. The additive variance exceeded the dominance variance in both contents, as the average degree of dominance was 0.58−0.88. Most of the dominance effects reduced these contents. The sense heritability was high, ranging from 0.859 to 0.903. For the sesamin and sesamolin contents, no major differences were observed between the F_1 and F_2 progeny. There was a statistically significant and positive correlation with each other in each generation. Thus the content of sesamin and sesamolin would be controlled by a common polygenic system. Selecting new sesame lines with high sesamin and sesamolin contents based on the phenotype of low generations should be effective.

4.6.3 Seed Coat Type

Seed coat type (including seed coat color, roughness, and hardness) affects the commodity quality of sesame. The gene interaction of seed coat color (from white to black), roughness (from rough to smooth), and hardness (from stiff to tender) traits have been studied in sesame. The

results showed that the above two traits were under the control of single gene. As to the seed coat color, the inheritance is complex (Culp 1959; Shahidi et al. 2006; El-Bramawy and Shaban 2008; Kanu 2011; Zhang et al. 2012).

4.6.4 Seed Coat Hardness

The inheritance of seed coat hardness (stiffness or tenderness) in sesame has been studied since the early twentieth century (Nohara 1930). The stiffness of the seed coat of black-and brown-coat seeds is chiefly due to the thickening of the lower part of the side wall in the epidermal cells with which a deposition of something like the crystals of calcium oxalate on the bottom wall is completely linked. The tenderness of the white seeds is due to the natural wall of the epidermal cells contrary to the above stiff seed coat. The hereditary nature of stiffness to tenderness seed coat trait follows the simple Mendelian rule, as the stiffness is dominant over the tenderness with the heterozygotes segregation ratio (stiffness: tenderness = 3:1) in the F_2 generation.

4.6.5 Seed Coat Color

The two sesame lines (K8: smooth seed coat type; N124: rough seed coat type) used as parents in Culp (1959) study to investigate the inheritance of seed coat type in sesame. The F_1, and BC_1 ($F_1 \times$ N124) plants possessed rough seed, indicating dominance of this character. The F_2 generation of 252 plants consisted of 185 rough and 67 smooth seeded plants. This segregation conforms to an expected 3:1 ratio (Chi-square = 0.339). The BC_2 ($F_1 \times$ K8) of 112 plants was composed of 49 rough and 63 smooth seeded segregates. This conforms to an expected 1:1 ratio (Chi-square = 1.750). The above results indicate the presence of rough or smooth seed coat type is simply inherited and is conditioned by one pair of genes with rough seed coat dominant.

Seed coat color is a commercial trait in sesame. Seed coat color is also associated with antioxidant content and activity, seed biochemical properties, disease resistance level, and even genus evolution character (Budowski and Markley 1951; Nakimi 1995; Shahidi et al. 2006; El-Bramawy and Shaban 2008; Kanu 2011; Zhang et al. 2012, 2013). Thus, the inheritance of sesame seed coat color has been studied over a long period. In sesame, three indicators, L*, a*, and b* are used to exhibit the variance of the seed coat color trait. Of the 763 worldwide sesame accessions cultivated in 2017, the values of L*, a*, and b* ranged from $15.17 \sim 67.21$, $-0.07 \sim 14.84$, $-0.5 \sim 30.29$, respectively (Unpublished data, Haiyang Zhang,).

Nohara (1933) primarily performed a cross between a white-seeded and a black-seeded sesame accession, and found the ratio of black: dark brown: pale brown: white seed coat color accorded with 9:3:3:1. Seed coat color trait should be controlled by two genes. Sikka and Gupta (1947) also proved that black cost color is invariably dominant and seed coat color is regulated by two genes. Gutierrez et al. (1994) found that black is the dominant testa color, and light brown was observed to be partially dominant over white. The results showed that the seed coat color trait is controlled by two independent genes with complementary effects and complete dominance at each locus. However, genetic analysis results of the F_2 generation in black $\times$ light brown and black $\times$ white crosses revealed that one gene had complete dominance and supplemented the effects of other genes controlling basic testa colors. Baydar and Turgut (2000) reported that the colored seed was dominant over white seed. However, F_2 generations exhibited 3:1 segregation ratio, as well as 9:4:3 and 9:3:4 ratios, indicating the inheritance of both complete dominance and epistatic action types.

In addition, Falusi (2007) performed the genetic analysis of crosses from nine sesame accessions and demonstrated that seed color had a complex genetic basis. Different gene actions were responsible for the expression of the trait. Single gene action was observed in three crosses and two gene action in two crosses. In the sixth cross involving white and black seeded parents, an array of white, black, brownish-white, and

brown colors were observed in the F_2 and revealed the complex genetic nature. Results also indicated that plants with the same seed color might be under different genotype constitution. Pandey et al. (2013) studied the genetic base of sesame seed color using the five crosses derived from different varieties with white, beige, or various shades of brown color. In general, tetragenic model corroborated with the color combinations. The color beige seemed fixable as well as suppressive over other colors. In white seeded lines, several genes co-existed and produced segregants in F_2 generation. Two loci, *Gr* and *I,* regulated the suppression or intensification of pigment production. Two major genes, *V* and *B* were accountable for basic color production. The recessive allele '*b*' tightly linked with '*r*', resulted in beige color which suppressed the effect either of the loci *V* and *B*. In recent years, Zhang et al. (2013) analyzed genetic segregations of the six generations of a cross and performed the quantitative trait loci (QTL) for sesame seed coat color. The results showed that two major genes with additive-dominant-epistatic effects along with polygenes controlled the seed coat color trait in sesame.

4.7 Disease Resistance in Sesame

For sesame, the seed yield and quality are always affected by biotic (such as diseases and insects) and abiotic (such as waterlogging, drought, and wind) stresses (Rajput et al. 1998; Verma et al. 2005; Yuan et al. 2018). There are at least eight fungal diseases underlying the economical attack to sesame in the worldwide production areas (Kolte 1985). Especially, the three diseases, i.e., *Fusarium* wilt, charcoal rot, and *Alternaria* leaf spot is the most important diseases which usually cause about 25–40% of yield losses in sesame (El-Bramawy 2003, 2006; El-Bramawy and Shaban 2008). Enhancing the disease resistance of the host plants to pathogens is the principal approach to reduce the yield loss induced by diseases. Meanwhile, clarifying the inheritance pattern of disease resistance is prerequisite to realize the entire improvement of the disease resistance in sesame varieties.

4.7.1 *Fusarium* Wilt Disease Resistance

Sesame *Fusarium* wilt (SFW) disease is one worldwide disease in sesame and can cause 15–30% of the yield loss annually (Bedigian and Harlan 1986; Verma et al. 2005). SFW is induced by *Fusarium oxysporum* f. sp. *sesami* (FOS). SFW disease is primarily found in Egypt, India, the Soviet Union, and other sesame production countries since 1920s. Besides host genotypes and environmental factors, the *Fusarium* wilt resistance level of a sesame variety can be affected by pathogen races and pathogenicity (Qiu et al. 2014; Miao et al. 2019). In recent a few years, the Chinese sesame scientists systematically observed the SFW infection in hundreds of sesame germplasm accessions under the artificial nursery (Yuanyang experimental station, China) in successive two years (Fig. 4.1) and determined the *Fusarium* wilt infection scales according to the disease index (DI) (Qiu et al. 2014; Miao et al. 2019).

The disease index was calculated according to the following formula:

$$\mathrm{DI} = \frac{\sum (S_i \times n_i)}{2n} \times 100$$

DI: disease index; s_i: symptom grade; n_i: plants infected by *Fusarium* wilt pathogen; *N*: total plant number.

There are five grades from the highly resistant (HR) to highly susceptible (HS) with the successive DI value scales (Table 4.10).

For the two groups of the susceptible sesame varieties or lines inoculated with FOS strains in 2015 and 2017, the peak value of DI appeared at the mature stage, with a high DI of more than 55. However, the key stages leading to the DI increase were mainly from budding to early flowering stage (about in the middle of July) and the full flowering stage (from the middle to the

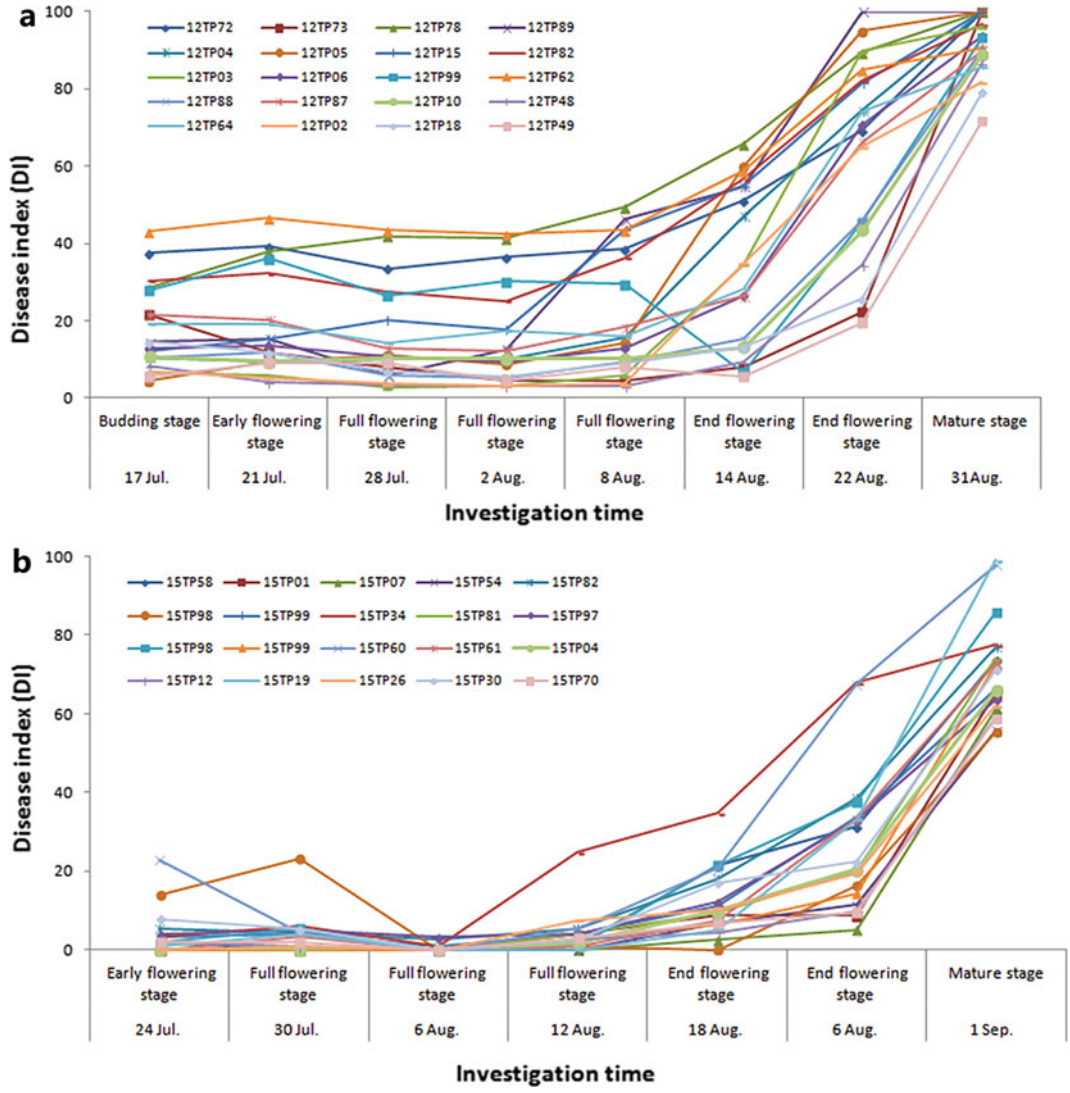


Fig. 4.1 Disease index (DI) variation of susceptible sesame varieties inoculated with FOS strains in nursery during two seasons. All the varieties were grown in the Yuanyang *Fusarium* wilt disease nursery (E 113.97º, N 35.05º). The nursery was inoculated with five FOS strains with high pathogenicity. **a**: DI variation of 20 varieties highly susceptible to *Fusarium* wilt grown in 2015. **b**: DI variation of 20 varieties highly susceptible to *Fusarium* wilt grown in 2017 (Provided by Hongmei Miao)

end of August) (Unpublished data, Haiyang Zhang) (Fig. 4.1).

As shown in Fig. 4.1, in each of the 20 sesame varieties with high susceptibility to FOS strains, the value of DI variation rapidly increased mainly at seedling and flowering stages (Unpublished data, Haiyang Zhang). In order to explore the genetic inheritance of sesame germplasm to FOS pathogen, the *Fusarium* wilt resistance of the 112 lines of $F_{2\text{-}3}$ progeny derived from the cross JS012 × DS899 and the parents were evaluated using the laboratory detection method described by Qiu et al. (2014). The results showed that two FW resistance QTLs

were determined with the high R^2 values of 0.11 and 0.15, respectively (Unpublished data, Haiyang Zhang).

In the past few decades, sesame researchers performed the *Fusarium* wilt resistance evaluation of sesame germplasm accessions and cross lines under the natural field or artificial nursery conditions and found that the SFW resistance trait was complex with different gene interaction effects in various cross combinations (Wang et al. 1993; El-Bramawy 2006). Wang et al. (1993) systematically investigated the *Fusarium* wilt resistance in F_1, F_2, and backcross generations of diallel crosses of five sesame varieties at full flowering stage. Analysis results indicated that the inheritance of the resistance to FOS was mainly controlled by additive gene effect. The GCA of disease index was the most significant in F_1 and F_2, while some crosses presented the gene interaction effects. Meanwhile, El-Bramawy (2006) observed two generations (F_3 and F_4) derived from the 15 crosses during two successive seasons. All the samples were cultured under natural inoculation of FOS pathogens. The results showed that the heritability for *Fusarium* wilt resistance in sesame was very high (more than 95%) in both generations. The screened lines with the highest significant correlation could be applied for new variety breeding (El-Bramawy 2006). Subsequently, El-Bramawy and Shaban (2008) investigated the *Fusarium* wilt resistance of 15 F_1 and six parents and detected that the infection percent of the 21 groups varied from 6.12% to 30.10%. Besides a predominance of dominance components, additive and non-additive components of genetic variance were found contributing to the inheritance behavior of the resistance. In addition, the additive, dominance, and epistatic components of genetic variation were also detected in some cross families (Bakheit et al. 2000).

4.7.2 Charcoal Stem Rot Resistance

Charcoal stem rot disease (CSR) is one of the most important worldwide diseases for sesame (Rajput et al. 1998; El-Bramawy, and Wahid 2006), as it is destructive in all stages of sesame cycle and can cause about 5–100% of the yield loss (Vyas 1981). CSR is caused by *Macrophomina phaseolina* (Tassi.) Goid. The most common symptom of the charcoal rot disease is the sudden wilting mainly at the flowering stage. The stem and root of the infected plants always get black rapidly due to severe tissue destruction by *Macrophomina* pathogen (Fig. 4.2).

As the infection of charcoal rot disease is related with the genotypes and environments, Zhang and Feng (2006) set up the charcoal rot resistance evaluation standard for sesame. The investigation time is the day by which over 50% plants in a population present the charcoal rot symptoms during flowering stage. The disease degree is designed as follows: 0: normal plant without disease spots; 1: less than 1/3 of the plant and less than 1/4 of the capsules on plant exhibit charcoal rot; 3 : 1/3 to 2/3 of the plant and 1/4 to 1/2 of the capsules on plant exhibit charcoal rot; 5: over 2/3 of the plant and 1/2 to 3/4 of the capsules on plant exhibit charcoal rot; and 7: plant dead (Zhang and Feng 2006; Wang et al.

Fig. 4.2 Symptoms of charcoal stem rot disease in sesame. Left image indicates the initial symptom. Right image indicates later symptom of charcoal rot disease. (Provided by Hongyan Liu)

2017). The disease index (DI) was calculated based on the following formula:

$$\mathrm{DIN} = \sum n_i = 1(X_i \times I) / \left(7 \times \sum n_i = 1X_i\right),\ i = 1, 3, 5, \text{and } 7$$

(where *Xi* represents the plant number with disease degree *I*).

Meanwhile, Thiyagu et al. (2007) established the artificial screening and sick plot methods to evaluate the resistance of sesame genotypes to charcoal rot disease. The resistance was classified into five grades with the corresponding infection percentages, i.e., resistant, moderately resistant, moderately susceptible, susceptible, and highly susceptible (Dinakaran and Mohammed 2001).

To reveal the genetic character of resistance to charcoal rot in sesame, El-Bramawy and Wahid (2006) analyzed the charcoal rot resistance of the F_3 and F_4 generations of 6 × 6 diallele crosses under natural infection conditions in 2004 and 2005. The results presented the significant differences of the charcoal rot resistance level in various crosses and the characters of quantitative trait in sesame. Regress analysis indicated that seed yield, branch number, and seed color were significantly correlated with the infection percent of charcoal rot and *Fusarium* wilt in sesame (El-Bramawy and Wahid 2006; El-Bramawy and Shaban 2008). Li and Zhang (1984) investigated the resistance to charcoal stem rot in F_1 progeny of 28 crosses and eight parents in the successive two years. The results reflected that some F_1 progeny presented heterosis in disease resistance. The resistance variation in different crosses was significant and the resistance to charcoal rot was controlled by dominant genes. Considering the heterosis and dominant gene effect, screening the elite homozygous lines with high disease resistance from the low generation progeny seems uneasy (Li and Zhang 1984). Similarly, Sinhamahapatra and Das (1992) analyzed the combining ability analysis of a 9 × 9 diallel set of sesame genotypes under artificial infection plot inoculated with *Macraphamina phasealina* pathogen. The results revealed that the dominance component played a major role in controlling the charcoal rot resistance in the population.

El-Bramawy and Shaban (2008) investigated the charcoal rot disease resistance of the 15 F_1 and 6 parents under greenhouse conditions according to the method of Ahmad et al. (2002). The charcoal rot infection ratio of the 21 groups of samples varied from 2.95 to 24.11%. Similar to FOS resistance, the inheritance behavior for charcoal rot resistance should be controlled by both additive and non-additive components with a preponderance of non-additive genetic variance. The parents possessed mostly negative genes in dominant form with respect to resistance to charcoal rot trait. Moreover, Pandey et al. (2017) prove that both dominant and additive gene actions were involved in the charcoal rot resistance in sesame, and the magnitude of dominance gene effects was high. The genetic analysis indicates that the parents with high resistance to charcoal rot could be used as elite materials for disease resistance breeding of new sesame varieties.

4.7.3 Leaf Disease Resistance

Leaf diseases are common diseases in sesame (Fig. 4.3). There are five main worldwide leaf diseases in sesame, i.e., sesame leaf blight caused by *Helminthosporium sesami* Miyake, Alternaria leaf spot (or black spot) caused by *Alternaria sesami* (Kawamura) Mohanty et Behera, Nigrospora leaf blight by *Nigrospora sphaerica*, brown spot by*Ascochyta sesami* Miura, and Corynespora blight (or Corynespora blight spot) by *Corynespora cassiicola* (Berk and Curt) (Fungi Identification Manual (First edition), Crop Diseases and Insect Pests in China (Second edition) (Naik et al. 2003; Sun 2006; Zhao et al. 2014). Independently, leaf blight disease can cause the infection of 30–40% of the population and always result in serious losses for sesame production (Zhao et al. 2014). Some leaf diseases such as *Alternaria* leaf spot can cause seed rot, pre-or post-emergence death of seedlings, infect all the above-ground parts, and result in considerable yield loss and seed quality (Naik et al. 2003).

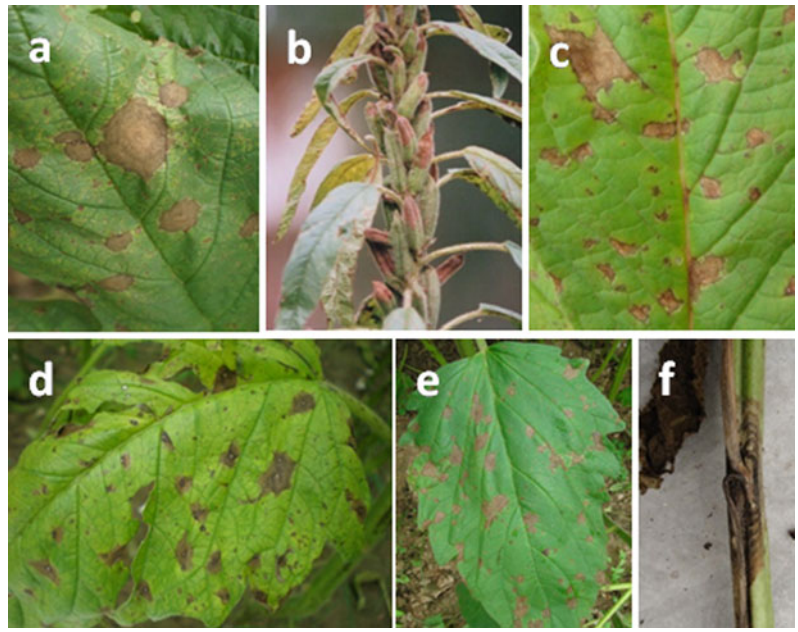


Fig. 4.3 Symptoms of sesame leaf diseases. **a**: Sesame leaf blight in leaf; **b**: Sesame leaf blight in capsule; **c**: Sesame *Alternaria* leaf spot in leaf; **d**: Sesame brown spot in leaf; **e**: Sesame *Corynespora* leaf spot in leaf; **f**: Sesame *Corynespora* leaf spot in stem (Provided by Hongyan Liu)

In sesame fields, two or more leaf diseases often occur simultaneously in a population. Thus, performing the genetic analysis of leaf disease seems more difficult. Till now only several leaf diseases are applied for the genetic inheritance in sesame. El-Bramawy and Shaban (2008) analyzed the *Alternaria* leaf spot resistance of 45 d-old sesame plants under greenhouse conditions in 2005 and 2006. All the seedlings were sprayed with *Alternaria* leaf spore suspension (2×10^3 spore per mL) and the disease syndrome was assessed 1 week after inoculation according to the method of Karunanithi (1996). The resistance to *Alternaria* leaf spot presented mostly additive variance. In order to explore the inheritance of *Alternaria* blight resistance in sesame, Eshwarappa (2010) investigated the cross RT-273 (resistant) × Gulbarga local black (susceptible) and determined that the resistance to *Alternaria* blight disease is controlled by single dominant gene. Testing analysis of the F_3 and F_4 families under field condition confirmed the single dominance of the gene action (Deepa 2010; Goudappagoudar et al. 2014).

4.7.4 Powdery Mildew Disease Resistance

Powdery mildew disease is caused by *Oidium acanthospermi* (Childarwar) and can result in 50% yield losses in sesame (Venkata et al. 2013). Powdery mildew mainly occurs in India and some countries with heavy rainfall environment. To precisely investigate the disease resistance, Rao et al. (2012) scored the disease intensity using the six grades (from 0 to 9) (Table 4.11). Based on the percentage disease index (PDI), the disease resistance is grouped into four categories, i.e., 0: indicates immune (I), 1–30: resistant (R), 31–50: moderately resistant (MR)/tolerant (T), and 51–100: susceptible (S) (Raja 1990).

The resistance level to powdery mildew disease is determined using the percentage disease index (PDI) according to the following formula described by McKinney (1923) (Rao et al. 2012).

$$\text{PID} = \frac{\text{Sum of grades}}{\text{Total number of leaves analysed} \times \text{Maximum disease grade}} \times 100$$

To efficiently evaluate the disease resistance, Venkata et al. (2013) established the powdery mildew disease screening method. In the early 1970s, scientists initiated cross combinations and resistance gene action analysis. Some reports showed that two major genes with complementary gene action were involved in the resistance to powdery mildew disease in sesame (Raja Ravindran and Rathinam 1996; Rao et al. 2012; Reddy and Haripriya 1990). Moreover, Rao et al. (2012) investigated the powdery mildew tolerance of the 6 generations derived from the cross PKDS 37 × Swethatil under natural field conditions. The results indicated that the resistance to powdery mildew was recessive and should be controlled by two independent recessive genes with complementary epistasis (Rao et al. 2012). Similarly, Sravani et al. (2012) reported that the resistance to powdery mildew disease was governed by two pairs of recessive genes. Venkata et al. (2013) found that the susceptibility was dominant over the resistance phenotype using new evaluation method. The different results might be resulted from the various resistant sources.

4.7.5 Phyllody Resistance

Phyllody is one of the most destructive diseases in sesame and causes serious yield losses. Phyllody is caused by a pleomorphic mycoplasma-like organism (phytoplasma) and could be transmitted by leaf hopper (Vasudeva and Sahambi 1955; Ustun et al. 2017). In general, the shoots of infected plants present shorten internode with high dense leaves, and plants grow stunted. Meanwhile, the floral organs keep green with leaf-like shape (Win et al. 2010). As a result, the infected plants hardly form normal flowers, capsules, and seeds and result in obvious yield loss.

Reliable screening methods for phytoplasma inoculation were described by Singh et al. (2007). Periwinkle was applied as an alternate host to transmit the disease to healthy sesame plants through dodder (*Cuscuta campestris*) transmission (Singh et al. 2007). Dodders were used as the bridge between the recipient host (healthy sesame) and the donor plant. As to the resistance inheritance to phyllody, the gene action type of the trait in sesame has been studied with so many populations and crosses. Most results showed that the resistance to phyllody is mainly controlled by one or two genes in sesame. For instance, Saravanan et al. (2003) reported that additive genetic variance was more important for phyllody occurrence. Naik et al. (2011) studied the inheritance pattern of phyllody resistance with the cross derived from a resistant landraces Tamil Nadu local (TNL) and a susceptible cultivar E8. The segregation analysis of a total of 149 F_3 families revealed that phyllody resistance is under the control of two dominant genes with complementary gene action (9:7). Singh et al. (2007) selected three intra- and two inter-specific crosses to study the inheritance of phyllody resistance, whereas two each intra- and interspecific crosses were identified to test the allelic relationship between resistant genes. The segregation analysis of F_2 and the backcross of the intraspecific crosses revealed that a single recessive gene controlled phyllody resistance, whereas the results of the interspecific combinations suggested the involvement of a single dominant gene. Allelic test on intraspecific crosses revealed the recessive resistance by two independent non-allelic genes exhibiting duplicate dominance, whereas the interspecific crosses presented the dominance nature of the resistance resulted from the involvement of one dominant and one recessive gene.

4.8 Insects Resistance in Sesame

Besides the above factors, insects are also a major limit to high and stable yield for sesame production. About 65 species of insects can

attack sesame at different growth and development stages (Ahuja and Bkhetia 1995).

4.8.1 Gall Midge Resistance

Sesame gall midge is caused by *Asphondylia sesami* Felt. To understand the genetic character and the resistance mode to gall midge, Ubor et al (2015) screened the gall midge resistance using 30 sesame genotypes and the crosses under field conditions. The result showed that non-additive gene action took the function in the inheritance of resistance to gall midge. Cross analysis indicated that the additive and non-additive gene actions simultaneously controlled the inheritance of the resistance. Estimation of the genetic effects for the resistance presented the predominance of additive and additive × additive type of epistasis, even though dominance also had a role in the cross Sesim1 × 7020-1-2.

4.8.2 Shoot Webber Resistance

Foliage feeder shoot webber (*Antigastra catalaunalis* D.) occurs widely in most sesame growing regions. Shoot webber can cause huge damage up to 40% of population in India especially in Tamil Nadu (Abraham et al. 1977). Gnanasekaran et al. (2010) analyzed the gene action of the resistance to shoot webber using eight varieties with high yield level and five genotypes with high resistance to shoot webber. The results indicated that the resistance to shoot webber in the samples was highly influenced by non-additive gene action.

References

Abraham EV, Natarajan K, Murugesan M (1977) Damage by pest phyllody to sesame in relation to time of sowing. Madras Agric J 64:293–301

Agrawal MM, Singh S, Wawge MN, Macwana S, Sasidharan N (2017) Correlation and path analysis for seed yield and yield attributing traits in Sesame germplasm (*Sesamum indicum* L.). Intl J Chem Stud 5:1099–1102

Ahmad R, Mahmood T, Saleem MF, Ahmad S (2002) Comparative performance of two Sesame (*Sesamum indicum* L.) varieties under different row Spacing. Asian J Plant Sci 1:546–547

Ahuja DB, Bkhetia DRC (1995) Bioecology and management of insect pests of sesame. A review. J Insect Sci 8:1–19

Akbar F, Rabbani MA, Shinwari KZ, Khan JS (2011) Genetic divergence in Sesame (*Sesamum indicum* L.) landraces based on qualitative and quantitative traits. Pak J Bot 43:2737–2744

Anuradha T (2002) Genetic divergence stability, heterosis and genetic analysis of yield and yield attributes in sesame (*Sesamum indicum* L.). PhD Thesis, Acharya N. G. Ranga Agricultural University, Hyderabad

Anuradha T, Lakshmi KRG (2008) Nature of gene action for quantitative traits in sesame. Agric Sci Digest 28:307–308

Armenda-Soto J, Chang TT, Loresto GC, O'Toole JC (1983) Genetic analysis of root characters in rice (*Oryza sativa* L.). Soc Adv Breed Asia Oceania 15:103–116

Ashri A (1981) Increased genetic variability for sesame improvement by hybridization and induced mutations. In: Ashri A (ed) Sesame: Status and Improvement. Food and Agriculture Organization of The United Nations, Rome, Italy, pp 25–27

Ashri A (1998) Sesame breeding. In: Janick J (ed) Plant breeding reviews, vol 16. Wiley, Israel, pp 179–228

Aye M, Win S, Hom N (2018) Combining ability and heterosis studies in sesame (*Sesamum indicum* L.) genotypes. Intl J Adv Res 6:1220–1229

Babu DR, Kumar PVR, Rani CVD, Reddy AV (2004) Studies on combining ability for yield and yield components in sesame (*Sesamum indicum* L.). J Oilseeds Res 21:260–262

Bakheit BR, Ismail AA, Elshiemy AA, Sedek FS (2000) Triple test cross analysis in four sesame crosses (*Sesamum indicum* L.) 2. yield, yield components and wilt infection. J Agric Sci 137:185–193

Bakheit BR, Ismail AA, Elshiemy AA, Sedek FS (2001) Triple test cross analysis in four sesame crosses (*Sesamum indicum* L.). J Agric Sci 137:185–193

Banerjee PP, Kole PC (2006) Genetic variability and yield analysis in sesame (*Sesamum indicum* L.). Crop Res 32:430–433

Banerjee PP, Kole PC (2009a) Analysis of genetic architecture for some physiological characters in sesame (*Sesamum indicum* L.). Euphytica 168:11–22

Banerjee PP, Kole PC (2009b) Combining Ability Analysis for Seed Yield and Some of its Component Characters in Sesame (*Sesamum indicum* L.). Intl J Plant Breed Genet 3:11–21

Baydar H, Turgut İ (2000) Studies on genetics and breeding in sesame (*Sesamum indicum* L.) I. Inheritance of the characters determining the plant type. Proceedings of SPIE—The International Society for Optical Engineering 6935:93–98

Bedigian D, Harlan JR (1986) Evidence for cultivation of sesame in the ancient world. Econ Bot 40:137–154

Biabani AR, Pakniyat H (2008) Evaluation of seed yield-related characters in sesame (*Sesamum indicum* L.) using factor and path analysis. Pak J Biol Sci 11:1157–1160

Budowski P, Markley KS (1951) The chemical and physiological properties of sesame oil. Chem Rev 48:125–151

Chandramony D, Nayer NK (1988) Diallel analysis in sesame (*Sesamum indicum* L.). Agri Sci Digest 8:193–198

Chaudhari GB, Anarase SA, Ban YG (2016) Combining ability and gene action over environments in Sesame (*Sesamum indicum* L.). Elect J Plant Breed 7:71–77

Chavan AA, Makne VG, Chopde PR (1981) Gene action in sesame. Indian J Genet 41:419–422

Chun JA, Jin UH, Lee JW, Yi YB, Hyung NI et al (2003) Isolation and characterization of a myo-inositol 1-phosphate synthase cDNA from developing sesame (*Sesamum indicum* L.) seeds: functional and differential expression, and salt–induced transcription during germination. Planta 216:874–880

Culp TW (1959) Inheritance and association of oil and protein content and seed coat type in sesame, *Sesamum indicum* L. Genetics 44:897–909

Daniya E, Dadari SA, Ndahi WB, Kuchinda NC, Babaji BA (2013) Correlation and path analysis between seed yield and some weed and quantitative components in two sesame (*Sesamum indicum* L.) varieties as influence by seed rate and nitrigen fertilizer. J Biol Agri Healthcare 3:12–16

Das S, Gupta TD (1999) Combining ability in sesame. Indian J Genet 59:69–75

Deepa (2010) Phenotyping and molecular characterization of F3 progenies for Alternaria blight resistance in sesame (*Sesamum indicum* L.). MSc (Ag) Thesis, University of Agricultural Sciences, Raichur, Karnataka, India

Deepa PS, Lokesha R (2012) Studies on general and specific combining ability studies in sesame (*Sesamum indicum* L.). Intl J Plant Sci 7:391–394

Dinakaran D, Mohammed N (2001) Identification of resistant sources to root rot of sesame caused by *Macrophomina phaseolina* (Tassi.) Goid. Sesame & Safflower Newsl 16:68–71

Duc Pham T, Thi Nguyen TD, Carlsson AS, Bui TM (2010) Morphological evaluation of sesame (*Sesamum indicum* L.) varieties from different origins. Aust J Crop Sci 4:498–504

Ebata Y, Electron T (2010) Assessment of selection criteria in sesame by using correlation coefficients, path and factor analyses. Aust J Crop Sci 4:598–602

El-Bramawy MA, Wahid OAA (2006) Field resistance of crosses of sesame (*Sesamum indicum* L.) to charcoal root rot caused by *Macrophomina phaseolina* (tassi.) Goid. Plant Protect Sci UZPI (Czech Republic) 42:66–72

El-Bramawy MAS (2003) Breeding studies for Fusarium wilt resistance in sesame (*Sesamum indicum* L.). PhD Thesis. Argon Suez Canal Univ, Ismailia, Egypt

El-Bramawy MAS (2006) Inheritance of resistance to fusarium wilt in some sesame crosses under field conditions. Plant Protect Sci—UZPI (Czech Republic) 3:99–105

El-Bramawy MAS, Shaban WI (2008) Short communication. Inheritance of yield, yield components and resistance to major diseases in *Sesamum indicum* L. Span J Agric Res 6:623–628

El-Satar MAA, Ahmed FHA, Elnenny EMM (2016) Line x tester analysis of yield and it's components for high plant density tolerance in sesame. Egypt J Plant Breed 20:1009–1034

Eshwarappa V (2010) F2 Bulk segregants analysis for Alternaria disease resistance using RAPD markers in sesame (*Sesamum indicum* L.). MSc (Ag) Thesis, University of Agricultural Sciences, Dharwad, Karnataka, India

Falusi OA (2000) Genetic studies in the genes 'Sesamum' (Pedaliaceae). PhD Thesis submitted to Department of Crop Production, Federal University of Technology, Minna, Nigeria

Falusi OA (2007) Segregation of genes controlling seed colour in sesame (*Sesamum indicum* linn.) from Nigeria. Afr J Biotechnol 6:2780–2783

Falusi O, Salako EA, Funmi FM (2002) Inheritance of hairiness of stem and petiole in a selection from local (Nigeria) germplasm of sesame. Tropicultura 20 (3):156–158

Fatteh UG, Shah RM, Bodar DG (1982) Studies in combining ability in sesame (*Sesamum indicum* L.). Madras Agric J 69:145–150

Gaikwad K (2004) Genetic architecture of yield and yield attributing characters in sesame (*Sesamum indicum* L.). Crop Improv 36:1–5

Gangadhara k, Chandra PJ, Badiger B, Shadakshari TV, Yathish KR et al. (2012) Genetic divergence, genetic advance and heritability in sesame (*Sesame indicum* L.). A Quart J Life Sci 9(4):457–462

Gawali CW, Bhoite KD, Pardeshi SR, Mhaske BM, Wagh MP (2007) Study of genetic variability and correlation studies in sesame (*Sesamum indicum* L.). Agric Sci Digest 27:282–284

Gelalcha S, Hanchinal RR (2013) Correlation and path analysis in yield and yield components in spring bread wheat (*Triticum aestivum* L.) genotypes under irrigated condition in Southern India. Afr J Agr Res 8:3186–3192

Gnanasekaran M, Jebaraj S, Muthuramu S (2008) Correlation and path-coefficient analysis in sesamum (*Sesamum indicum* L.). Plant Arch 8:167–169

Gnanasekaran M, Jebaraj S, Muthuramu S (2010) Breeding for seed yield and shoot webber (Antigastra catalaunalis D.) resistance in sesame (*Sesamum indicum* L.). Elect J Plant Breed 1:1270–1275

Goudappagoudar R, Lokesha R, Vanishree (2014) Inheritance of Alternaria blight resistance in sesame. Intl J Plant Sci 8(1):110–112

Goudappagoudra R, Lokesha R, Ranganatha AARG (2011) Trait association and path coefficient analysis for yield and yield attributing traits in sesame (*Sesamum indicum* L.). Elect J Plant Breed 2:448–452

Gutierrez E, Monteverde E, Quijada P (1994) Inheritance of seed coat color and number of locules per capsule in three cultivars of sesame *Sesamun indicum* L. Agron Trop 44:513–527

Haruna IM, Aliyu L, Olufajo OO, Odion EC (2012) Contributions of some growth characters to seed yield of sesame (*Sesamum indicum* L.). SABB J Food Agric Sci 2:9–14

Hossain S, Ford R, Mcneil D, Pittock C, Panozzo JF (2010) Inheritance of seed size in chickpea (Cicer arietinum L.) and identification of QTL based on 100-seed weight and seed size index. Aust J Crop Sci 4:126–135

Ibrahim SE, Khidir MO (2012) Genotypic correlation and path coefficient analysis of yield and some yield components in sesame (*Sesamum indicum* L.). Intl J Agrisci 2:766–773

Jatothu L, Dangi KS, Kumar SS (2013) Evaluation of sesame crosses for heterosis of yield and yield attributing traits. J Trop Agri 51(1–2):84–91

Jin UH, Lee JW, Chung YS, Lee JH, Yi YB et al (2001) Characterization and temporal expression of a ω-6 fatty acid desaturase cDNA from sesame (*Sesamum indicum* L.) seeds. Plant Sci 161:935–941

Joshi AB (1961) Sesamum. Indian Central Oilseed Committee Hyderabad, India, pp 31–36

Kamala T (1999) Gene action for seeds yield and yield components in sesame (*Sesamum indicum* L). Indian J Agric Sci 69:773–774

Kanu PJ (2011) Biochemical analysis of black and white Sesame seeds from China. Amer J Biochem Mol Biol 1:145–157

Karunanithi K (1996) Efficacy of fungicides in the control of powdery mildew of sesame caused by *Oidium acanthospermi*. Indian J Mycol Plant Pathol 26:229–230

Kolte SJ (1985) Diseases of Annual Edible Oil Seed Crops. Vol. II: Rapeseed, Mustard, Safflower and Sesame Diseases. CRC Press Inc, Boca Raton, FL

Krishna DM, Thirugnana SK, Ganesan J (2002) Combining ability and heterosis for reproductive efficiency in sesame (*Sesamum indicum* L.). Sesame & Safflower Newsletter 17:5–9

Kumar KB, Vivekanandan P (2009) Studies on combining ability studies in sesame (*Sesamum indicum* L.). Elect J Plant Breed 1:33–36

Kumar P, Madhusudan K, Nadaf HL, Patil RK, Deshpande SK (2012) Combining ability and gene action studies in inter-mutant hybrids of sesame (*Sesamum indicum* L.). Karnataka J Agric Sci 25:1–4

Kumar PS, Ganesan J (2004) Generation mean analysis in sesame (S*esamum indicum* L.). Indian J Agric Sci 38:227–230

Kumhar SR, Solanki ZS, Choudhary BR (2008) Studies on Genetic Variability, Character Association and Path Coefficient Analysis in Sesame (*Sesamum indicum* L.). Indian J Plant Genet Res 21:90–92

Langham DG (1944) Natural and controlled pollination in sesame. J Hered 35:255–256

Langham DG (1945a) Genetics of sesame. J Hered 36:135–142

Langham DG (1945b) Genetics of sesame II. Inheritance of seed pod number, aphid resistance, "yellow-leaf" and wrinkled leaves. J Hered 36:245–253

Langham DG (1946) Genetics of sesame III. Open sesame and mottled leaf. J Hered 37:149–152

Langham DG (1947) Genetics of sesame V. Some morphological differences of the sesame flower (*Sesamum indicum* L.). J Hered 38:347–352

Langham DR (2007) Phenology of sesame. In: Janick J, Whipkey A (eds) Issues in New Crops and New Uses. ASHS Press, Alexandria, VA, pp 144–182

Langham DR (2008) Growth and development of sesame. American Sesame Grower Association, San Antonio, TX, p 44

Langham DR (2017) VIII Capsule zone descriptors of sesame (*Sesamum indicum* L.). ResearchGate, p 276

Langham DR (2018) X Sesame seed descriptors (*Sesamum indicum* L.), Working paper 1. ResearchGate, p 141

Leduc V, Moneret-Vautrin DA, Tzen JTC, Morisset M, Guerin L et al (2006) Identification of oleosins as major allergens in sesame seed allergic patients. Allergy 61:349–356

Li C, Miao H, Wei L, Zhang T, Han X, et al. (2014) Association mapping of seed oil and protein content in *Sesamum indicum* L. using SSR markers. PLoS One 9:e105757

Li GS, Zhang YZ (1984) The genetic study of the disease resistance to charcoal rot in early sesame generations. Chin J Oil Crop Sci 1:12–17

Liu JR, Ding FK, Tu LC (1980) Association analysis of seed yield component characters in sesame (*Sesamum indicum* L.). Chin J Oil Crop Sci 2:10

Magni C, Ballabio C, Restani P, Fuggetta D, Alessandri C et al. (2010) Molecular insight into IgE-mediated reactions to sesame (*Sesamum indicum* L.) seed proteins. Ann Allerg Asthma Immumol 105(6):458–464

Mansouri S, Ahmadi MR (1998) Study of combining ability and gene effect on sesame lines by diallel cross method. Iran J Agric Sci 29:47–55

Mcharo T (1994) Diallel analysis of metric traits in sesame (*Sesamum indicum* L.). Msc Thesis, Department of Plant Science and Crop Protection, University of Nairobi, Kenya

Mckinney HK (1923) Influence of soil temperature and moisture on infection of wheat seedlings by Helminthosporium sativum. J Agric Res 26:195–218

Mei H, Liu Y, Du Z, Ke W, Cui C et al (2017) High-density genetic map construction and gene mapping of basal branching habit and flowers per leaf axil in sesame. Front Plant Sci 8:636

Miao HM, Li C, Duan YH, Wei LB, Ju M et al (2019) Identification of a Sidwf1 gene controlling short internode length trait in the sesame dwarf mutant dw607. Theor Appl Genet 133:73–86

Monpara B, Khairnar S (2016) Heritability and predicted gain from selection in components of crop duration and seed yield in sesame (*Sesamum indicum* L.). Plant Gene Trait 7:1–5

Monpara BA (2009) Relationship of durum wheat yield to agronomical and physiological growth parameters. Intl J Agric Sci 5:399

Mothilal A, Manoharan V (2005) Diallel analysis for the estimation of genetic parameters in sesame (*Sesamum indicum* L.). Agric Sci Digest 25:133–135

Muhamman MA, Mohammed SG, Lado A, Betel MD (2010) Interrelationship and path coefficient analysis of some growth and yield characteristics in sesame (*Sesamurn Indicum* L.). J Agri Sci 2:100–105

Mungala RA (2017) Study of combining ability for seed yield and its component in sesame (*Sesamum indicum* L.). Intl J Pure Appl Biosci 5:775–785

Murty DS (1975) Heterosis, combining ability and reciprocal effects for agronomic and chemical characters in sesame. Theor Appl Genet 45:294–299

Naik MK, Patil RG, Mestha RK, Sunkad G (2003) Epidemiology of alternaria blight of sesame. National seminar on stress management in oil seed for self reliance in vegetable oils. Indian Society of Oilseeds Research, DOR, Hyderabad, 28–30 January 2003

Naik MK, Shindle GG, Lokesha R, Ranganath AGR (2011) Inheritance study on phyllody resistance in sesame (*Sesamum indicum* L.). Plant Arch 11:777–778

Nakimi M (1995) The chemistry and physiological functions of sesame. Food Rev Intl 11:281–329

Nohara S (1930) Some characters of *Sesamum indicum*, l. and their inheritance. Jpn J Genet 6:180–185

Nohara S (1933) Genetic studies of *Sesamum indicum* L. J Cell Agric Tokyo Imperial Univ 12:227–386

Ogata N, Kato M (2016) Alf-diallel analysis for sesamin and sesamolin contents of sesame (*Sesamum indicum* L.) seeds. Jpn J Crop Sci 85:302–308

Padmasundari M, Kamala T, Rao YV (2010) Genetics of locule in *Sesamum indicum* L. Asian J Agric Sci 2:164–167

Pandey S, Chakraborty R, Dasgupta T (2017) Generation mean analysis to determine and partition the components of genetic resistance to root rot in sesame (*Sesamum indicum* L.). Indian Agric 60:34

Pandey SK, Das A, Dasgupta T (2013) Genetics of seed coat color in sesame (*Sesamum indicum* L.). Afr J Biotechnol 12:6061–6067

Passioura JB (1982) The role of root system characteristics in the drought resistance for crop plants. Drought Resistance in Cereals Crops with Emphasis on Rice 71–82

Pathak HC, Dixit SK (1988) Genetic analysis for single stemmed sesame (*Sesamum indicum* L.). Indian J Genet 48:325–330

Qiu CP, Zhang HY, Chang S, Wei LB, Miao HM (2014) Laboratory detecting method for pathogenicity of *Fusarium oxysporum* Schl. f. sp. *sesami* isolates. Acta Phytopathol Sin 44:26–35

Raikwar RS (2018) Diallel crossing for assessment of yield and its components in sesame (*Sesamum indicum* L.). Intl J Chem Stud 6:179–183

Raja Ravindran G, Rathinam AA (1996) Inheritance of resistance to powdery mildew in sesame. Sesame & Safflower Newsletter 11:82–84

Raja RG (1990) Genetics of powdery mildew resistance in sesame (*Sesamum indicum* L.). MSc (Agri) Thesis, TNAU, Coimbatore, India

Rajaravindran G, Kingshlin M, Shunmugavalli N (2000) Combining ability analysis in sesame (*Sesamum indicum* L.). Rese Crops 1:235–238

Rajput MA, Khan ZH, Jafri KA, Fazal AJ (1998) Field screening of sesame germplasm for resistance against charcoal rot (*Macrophomina phaseolina*). Sesame & Safflower Newsletter 13:63–66

Ramalingam A, Muralidharan V, Sheriff NM (1990) Combining ability studies in sesame. J Oilseeds Res 7:75–77

Ramesh Shekhawat N, Macwana SS, Choudhary R, Patel BR (2014) Line × tester analysis in sesame (*Sesamum indicum* L.). Bioscan 9:1657–1660

Rao PVR, Anuradha G, Jaya Prada M, Gouri Shankar V, Raja Reddy K et al (2012) Inheritance of powdery mildew tolerance in sesame. Arch Phytopathol Plant Protect 45:404–412

Reddy CD, Haripriya S (1990) Heterosis for tolerance of Powdery mildew in sesame. Indian J Mycol Plant Pathol 20:160–161

Sankar PD, Kumar CRA (2003) Combining ability studies on economic traits in sesame (*Sesamum indicum* L.). Indian J Agric Res 37:223–226

Saravanan S, Nadarajan N, Fernándezmartínez J (2003) Combining ability studies in sesame. Sesame & Safflower Newsletter 18:1–6

Saravanan T, Thirugnana KS, Ganesan J, Fernández-martínez J (2000) Combining ability and heterosis for earliness characters in sesame (*Sesamum indicum* L.). Sesame & Safflower Newsletter 15:7–13

Satish RG (2013) Genetic analysis of sesame (*Sesamum indicum* L.) for traits related to moisture stress tolerance with reference to root traits. Ph. D, Theses, University of Agricultural Sciences, Bengaluru

Saxena K, Bisen R (2017) Genetic Variability, Heritability and Genetic Advance for the Phenotypic Traits in Sesame (*Sesamum indicum* L.). Intl J Pure Appl Biosci 5:1126–1131

Shahidi F, Liyana-Pathirana CM, Wall DS (2006) Antioxidant activity of white and black sesame seeds and their hull fractions. Food Chem 99:478–483

Sharmila VG, Mahalingam (2007) Generation mean analysis for quantitative traits in sesame (*Sesamum indicum* L.) crosses. Genet Mol Biol 30:80–84

Shashidhar HE, Hittalmani S, Shivashankar G (1990) Heritability for root and shoot characters in drought

tolerant rice cultures. Current Research—University of Agricultural Sciences (Bangalore), pp 200–201
Sikka S, Gupta N (1947) Inheritance studies in *Sesamum orientale* L. Indian J Genet Breed 7:35–52
Singh AK, Singh SB, Singh AP, Sharma AK (2012) Genetic variability, character association and path analysis for seed yield and its component characters in wheat (*Triticum aestivum* L) under rainfed environment. Indian J Agric Res 46:53
Singh PK, Akram M, Vajpeyi M, Srivastava RL, Kumar K et al (2007) Screening and development of resistant sesame varieties against phytoplasma. Bull Insectol 60:303–304
Sinhamahapatra SP, Das SN (1992) Combining ability for resistance to charcoal rot (*macrophomina phaseolina* Tassi (Goid)) in sesame. Indian J Genet 52:261–263
Sravani D, Vijay Y, Bharathi D, Reedy V, Rao MVB et al. (2012) Genetics of powdery mildew resistance in sesame (*Sesamum indicum* L.). J Res ANGRAU 40 (1):73–74
Sumathi P, Muralidharah V (2010a) Analysis of genetic variability, association and path analysis in the hybrids of sesame (*Sesamum indicum* L.) Trop Agric Res Exten 13
Sumathi P, Muralidharan V (2008) Study of gene action and heterosis in monostem/shybranching genotypes in sesame (*Sesamum indicum* L.). Indian J Genet 68:269–274
Sumathi P, Muralidharan V (2009) Genetic analysis of the hybrids, involving monostem/shy branching in sesame (*Sesamum indicum* L.). Elect J Plant Breed 1:65–69
Sumathi P, Muralidharan V (2010) Inheritance of branching and important biometrical traits in sesame (*Sesamum indicum* L). Indian J Genet 70:97–101
Sumathi P, Muralidharan V (2014) Gene effects and inheritance of branching and other yield attributing characteristics in sesame (*Sesamum indicum* L). Trop Agric Res Exten 16:92–101
Sun X (2006) The Methodological Study on Taxonomy of the Genus *Alternaria* Nees. Ph. D, Thesis, Shandong Agricultural University, Taian, China
Tabatabaei I, Pazouki L, Bihamata MR, Mansoori S, Niinemets I (2011) Genetic variation among Iranian sesame (*Sesamum indicum* L.) accessions vis-a-vis exotic genotypes on the basis of morpho-physiological traits and RAPD markers. Aust J Crop Sci 5:1396–1407
Tanesaka E, Umeda E, Yamamoto M, Masuda K, Yamada K et al (2012) Inheritance mode of seed dormancy in the hybrid progeny of sesame, *Sesamum indicum*, and its wild relative. Sesamum mulayanum Nair. Weed Biol Manag 12(2):91–97
Thiyagu K, Kandasamy G, Manivannan N, Muralidharan V, Manoranjitham SK (2007) Identification of resistant genotypes to root rot disease (*Macrophomina phaseolina*) of sesame (*Sesamum indicum* L.). Agric Sci Digest 27:34–37
Ubor W, Gibson P, Anyanga W, Rubaihayo P (2015) Inheritance of resistance to sesame gall midge in uganda. Afr Crop Sci J 23:355–363
Ustun R, Yol E, Ikten C, Catal M, Uzun B (2017) Screening, selection and real-time qPCR validation for phytoplasma resistance in sesame (*Sesamum indicum* L.). Euphytica 213:159
Van Rheenen HA (1970) Intergeneric hybridization between *Ceratotheca sesamoides*, Endl. and *Sesamun indicum* L. Niger J Sci 4:251–254
Vasudeva RS, Sahambi HS (1955) Phyllody in sesame (*S. orientale* L.). Indian Phytopathol 8:124–129
Venkata P, Rao R, Anuradha G, Prasuna K, Shankar VG et al. (2013) Inheritance of powdery mildew resistance in sesame (*Sesamum indicum* L.)—a review. Intl J Bio-Resour Stress Manag 4(4): 614–619
Verma ML, Mehta N, Sangwan MS (2005) Diseases of oilseed crops. In: Saharan GS, Naresh M, Sangwan MS (eds) Fungal and bacterial diseases of sesame. Indus Publishing Company, New Delhi, pp 269–303
Vidhyavathi (2002) Study of heterosis and combining ability in sesame (*Sesamum indicum* L.) MSc (Ag) Thesis, Tamil Nadu Agricultural University, Coimbatore, India
Vidhyavathi R, Manivannan N, Muralidharan V (2005) Line x tester analysis in sesame (*Sesamum indicum* L.). Indian J Agric Res 39:225–228
Vyas SC (1981) Diseases of Sesamum and niger in India and their control. Pesticides 15:10
Wang L, Zhang Y, Li P, Wang X, Wen Z et al (2012) HPLC analysis of seed sesamin and sesamolin variation in a sesame germplasm collection in China. J Amer Oil Chem Soc 89:1011–1020
Wang L, Zhang Y, Zhu X, Zhu X, Li D et al (2017) Development of an SSR-based genetic map in sesame and identification of quantitative trait loci associated with charcoal rot resistance. Sci Rep 7:8349
Wang W, Liu J, Tu L (1993) Study on the inheritance of resistance to Fusarium wilt in sesame. Acta Agriculturae Universitatis Henannensis 21:84–89
Wei W, Zhang Y, Lv H, Li D, Wang L et al (2013) Association analysis for quality traits in a diverse panel of Chinese sesame (*Sesamum indicum* L.) germplasm. J Integr Plant Biol 55:745–758
Wei X, Liu K, Zhang Y, Feng Q, Wang L et al (2015) Genetic discovery for oil production and quality in sesame. Nat Commun 6:8609
Win NKK, Back CG, Jung HY (2010) Phyllody phytoplasma infecting sesame (*Sesamum indicum*) in Myanmar. Trop Plant Pathol 35:310–313
Wu K, Liu H, Yang M, Tao Y, Ma H et al. (2014) High-density genetic map construction and QTLs analysis of grain yield-related traits in Sesame (*Sesamum indicum* L.) based on RAD-Seq techonology. BMC Plant Biol 14:274
Yamanura KM, Nadaf HL (2009) Combining ability and gene action for yield and yield components in sesame (*Sesamum indicum* L.). Karnat J Agric Sci 22:255–260
Yermanos DM, Hemstreet S, Saleeb W, Huszar CK (1972) Oil content and composition of the seed in the world collection of sesame introductions. J Amer Oil Chem Soc 49:20–23

Yol E, Toker C, Uzun B (2017) Inheritance of long and dense capsule characteristics in sesame. Turk J Field Crops 22:8–13

Yol E, Uzun B (2011) Inheritance of number of capsules per leaf axil and hairiness on stem, leaf and capsule of sesame (*Sesamum indicum* L.). Aust J Crop Sci 5:78–81

Yuan QL, Zhang HY, Miao HM, Duan YH, Wei QC et al (2018) Effects of waterlogging stress on the quality of sesame seed and oil product. Acta Agri Bor-Sin 33:202–208

Zhang H, Miao H, Li C, Wei L, Duan Y et al (2016) Ultra-dense SNP genetic map construction and identification of SiDt gene controlling the determinate growth habit in *Sesamum indicum* L. Sci Rep 6:31556

Zhang HY, Miao HM, Li C, Wei LB, Ma Q (2012) Analysis of sesame karyotype and resemblance-near coefficient. Chin Bull Bot 47:602–614

Zhang HY, Miao HM, Ju M (2019) Potential for adaptation to climate change through genomic breeding in sesame. In: Kole C (ed) Genomic designing of climate-smart oilseed crops. Springer, Cham, Switzerland, pp 374–376

Zhang HY, Miao HM, Wei LB, Li CH, Zhao RH et al. (2013) Genetic analysis and QTL mapping of seed coat color in sesame (*Sesamum indicum* L.). PLoS One 8:e63898

Zhang X, Feng X (2006) Descriptors and Data Standard for Sesame (*Sesamum indicum* L.). China Agriculture Press, China

Zhao H, Liu HY, Yang XS, Liu YX, Tang L (2014) First report of Nigrospora leaf blight on sesame caused by *Nigrospora sphaerica* in China. Plant Dis 98:842

Mutagenesis for Creation of Genetic Variability in Sesame

5

Ming Ju, Hongmei Miao, Huili Wang, and Haiyang Zhang

Abstract

There are about 20,000 sesame germplasm accessions in the world. South Korea, India, China, and some African countries are the main countries reserving sesame germplasm. However, sesame is the sole cultivated *Sesamum* species for seed production. The relatively narrow genetic base limits the progress of the world's sesame breeding. Deficiency of the elite accessions also hinders the variety improvement. In the past decades, mutagenesis techniques have been applied for creation of new sesame germplasm and for breeding varieties. In this section, we summarize the main progress in mutation induction methods and mutant creation in sesame. Some key mutants with determinacy, dwarf, and other elite traits are enumerated in detail. The importance of mutagenesis approach in sesame is also discussed.

5.1 Introduction

At present, there are about 20,000 sesame germplasm accessions in the world. Most of the sesame germplasm materials are preserved in South Korea, India, China, and some African countries. Of which, 6658 accessions are preserved by National Bureau of Plant Genetic Resources (NBPGR) in India, 7698 accessions in South Korea, and more than 6000 accessions in China (Pei et al. 2014; Gong et al. 2016). Meanwhile, the USA, Japan, Venezula, Kenya, Senegal, Tanzania, and several other African countries also preserve some sesame germplasm accessions including some wild species (Personal communication, H. Zhang). In order to enhance the genetics and breeding research in sesame, the Chinese scientists initiated sesame germplasm collection and creation of genetic variability since 1950s. Till now, China has become one of the few countries constructing the sesame core germplasm population. Abundant germplasm resources provide the material basis for improving the genetics and breeding research in sesame. However, we have to keep in mind that sesame (*S. indicum*) is the sole *Sesamum* species that are directly cultivated for seed production. The relatively narrow genetic base limits the progress of the world sesame breeding. Especially, elite accessions with immunity to biotic stresses (including *Fusarium* wilt, Charcoal rot, and other major diseases), or significantly high resistance

M. Ju · H. Miao · H. Wang · H. Zhang (✉)
Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, People's Republic of China
e-mail: zhanghaiyang@zzu.edu.cn

H. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-98098-0_5

or tolerance to capsule shattering and abiotic stresses (including waterlogging, low temperature, and other abiotic stresses) are hardly detected from the naturally cultivated sesame reservoirs (Langham 2007; Qiu et al. 2014; Zhang et al. 2019). Therefore, creating new lines with desirable traits is a prerequisite for sesame breeding. Besides pedigree selection, mutation induction is the most effective technique for sesame variety improvement (Ashri 2001). Here we summarize progress in mutagenesis and thereby creation of new sesame germplasm accessions. Description of some key mutants with determinacy, dwarf, and other elite traits are also introduced.

5.2 Germplasm Creation Techniques

For sesame, the common mutagenesis method to induce mutants includes physical mutagenesis (such as X-ray, γ-ray, and fast neutron) and chemical mutagenesis (such as ethylmethane sulphonate (EMS), sodium azide, diethyl sulphate, and colchicine) (Ashri 1998; Muhammad et al. 2013). Scientists initiated physical and chemical mutagenesis research in sesame in the 1950s (Shi 1991). Most studies mainly focused on the optimization of the mutagen dosage and the treat time to determine the high efficiency of the mutagenesis techniques. Subsequently, space mutation, ion implantation, and hybridization breeding methods were gradually introduced into sesame breeding. In the 1990s, the Plant Breeding and Genetics Section (PBGS) of the Joint FAO/IAEA Division provided the coordinated research project (CRP) and stimulated the mutation induction and variety improvement in sesame (Zhang et al. 2019). As a result, worldwide sesame scientists from South Korea, China, Egypt, India, Japan, Israel, and several other countries performed the mutagenesis research in sesame. About 140 sesame mutants involved in several important agronomic traits and the disease resistances were created and reported (Ashri 1998, 2001; Van Zanten 2001; Kumari et al. 2016). In the past two decades, more mutants with short internode length and plant height, determinate inflorescence, high oil content, and other key elite traits were created and utilized for sesame variety improvement (Ramadoss et al. 2014; Begum and Dasgupta 2015; Zhang et al. 2016; Wang et al. 2017; Miao et al. 2020).

5.2.1 Physical Mutagenesis

Physical mutagens such as ultraviolet, X-ray, γ-ray, fast neutron, laser, microwave, and ion beam are effective reagents to induce new lines (Ashri 1998, 2001; Melzer et al. 2008). Especially X-rays, gamma rays, and fast neutrons are very common for sesame (Shi 1991; Uzun and Çağırgan 2009; Diouf et al. 2010). For example, the famous *dt*2-typemutant with determinate inflorescence was induced by γ-ray (5000 Grey) (Ashri 1998). Differing from another determinate mutant DS899 (*dt*1 type) induced by EMS mutagenesis, the *dt*2 mutation-induced by γ-ray resulted from the loss of a long DNA sequence containing the *SiDt* gene in sesame (Zhang et al. 2016).

Different sesame accessions present various resistance levels to mutagens to some extent. Pathirana (1994) found that the seeds of variety MI3 were more resistant to irradiation than MI2. Ramadoss et al. (2014) investigated the mutation efficiency of two genotypes TTVS 51 and TTVS 19 mutated by γ-rays and found that the high frequency of viable mutants was observed in the M_2 population with 450 Gy inducement condition. The visible mutation was involved in branching habit, plant height, phyllotaxy, the height to the first capsule, flower morphology, internode length, seed characters, and other agronomic traits. Further heritability studies confirmed that the M_3 bulk population of the radiated seeds induced under low dose (such as 250 Gy and 350 Gy) seemed more effective for mutant screening (Ramadoss et al. 2014).

As to fast neutron irradiation (FNI), Murty and Bhatia (1990) obtained a mutant (N29) with polypetalous corolla and no corolla lobes from the M_2 generation of CV.N.62–32 induced with 1.6 KR fast neutrons. Muhammad et al. (2013)

compared the tolerance level of the two cultivars Ex-Sudan and Kenana-4 with a flux of 1.5 × 104 ncm^{-2} s^{-1}. The results suggested that 12 and 16 μSv were the most effective dose to induce mutants with high yield potential.

5.2.2 Chemical Mutation Techniques

For sesame, the main chemical mutagens include EMS, sodium azide, diethyl sulphate, and colchicines (Shi 1991; Wang et al. 2017). Of which EMS is the most popular and widely applied for in sesame for the high convenience and efficiency (Kumar and Yadav 2010; Anbarasan et al. 2013, 2014, 2015; Wang et al. 2017).

Systematical research of EMS mutagenesis technique in sesame was initiated since the late twentieth century (Shi 1991; Begum and Dasgupta 2010; Lv et al. 2011; Kumar et al. 2012; Anbarasan et al. 2013; Kumari et al. 2016; Wang et al. 2017). Anbarasan et al. (2013) treated the sesame seeds with various EMS concentrations ranging from 0.2 to 2.0 mM. The mutation efficiency was evaluated based on the seed germination rate (%) of the 5th and 10th-day seedlings. The seedling height (shoot length), root length, and lateral branch number per primary root were also investigated and compared. As a result, 1.0 mM EMS touched the LD50 value (the 50% reduction of seed germination and seedling growth). In order to determine the optimal EMS concentration and the highest mutation efficiency in sesame, Wang et al. (2017) compared the mutation induction effects of the 15 treatment combinations with various EMS concentrations (0.1–2.0%) and treatment time (6–24 h). The results showed that the four treatment conditions (0.5%/24 h, 1.0%/24 h, 1.0%/12 h, and 1.5%/6 h) are the most suitable conditions for sesame. The mutation frequency touched to 0.80% or higher. For M_2 and M_3 generations of the visible mutants, 23.71% mutation could be stably descended to the progenies (Wang et al. 2017).

In order to induce the genetic variation of growth and yield traits in sesame, Salisu et al. (2011) performed chemical mutagenesis through colchicine treatments (including 0.1, 0.5, 1.0, and 2.0 mM) on the seeds of two varieties (var. Ex-Sudan and E-8). Some agronomic traits including germination percent, leaf number per plant, internode length, leaf area, plant height at maturity, and several yield-related traits were significantly and negatively affected ($P \leq 0.01$) with the increase of colchicine concentration. Meanwhile, some mutants with chlorophyll deficiency were also detected in the mutant population. The results suggested that low concentration of colchicine was suitable to induce the genetic variation of yield-related traits in sesame.

In recent years, mutagenesis studies mainly focused on the comparison analyses of various mutagens and mutation approach combinations (Begum and Dasgupta 2010; Boranayaka et al. 2010; Ravichandran and Jayakumar 2015). Especially, the effects of γ-rays and EMS with different concentration combinations on sesame were systematically investigated (Begum and Dasgupta 2010; Anbarasan et al. 2015). Comparison analysis results proved that the inducement efficiency of EMS (0.5, 1.0, 1.5, and 2.0%) was much higher than that of γ-rays (200, 400, and 600 Gy). The lowest concentration of γ-rays (200 Gy) and EMS (0.5%) could induce the highest mutation efficiency in the studied genotypes (Begum and Dasgupta 2010). Meanwhile, Boranayaka et al. (2010) performed the treatments of γ-rays of ^{60}Co (10, 20, 30, 40, and 50 krad) followed by EMS treatment (0.8, 1.0, 1.2, 1.4, and 1.6%) on seeds of two genotypes. The reduction of seed germination rate was more evident than other traits. Begum and Dasgupta (2011) assessed the extents of variability and heritability of the seed yield trait and yield components in the mutant populations of three sesame genotypes induced by γ-rays and EMS, respectively. The results reflected that lower doses of the physical and the chemical mutagens were more effective for mutation induction in sesame. EMS was more suitable to produce polygenic variability than γ-rays. 0.5% EMS appeared to be the best for creating new lines with high yield potential in early generations. Similarly, Ravichandran and Jayakumar (2015) studied the mutation efficiency of γ-rays and

EMS and proved that γ-irradiation was more effective to induce mutation than EMS treatment in sesame. Moreover, Anbarasan et al. (2015) compared the effects and the mutation efficiency of EMS (0.6, 0.8, 1.0 1.2, and 1.4 mM), γ-rays (30, 40, 50, 60, and 70 KR), and their combinations (30 KR + 0.6 mM, 40 KR + 0.8 mM, 50 KR + 1.0 mM, 60 KR + 1.2 mM, and 70 KR + 1.4 mM) on M_1 and M_2 mutant generations of sesame variety TMV3. In M_1 generation, the reduction of seed germination, survivability, and seedling height traits was observed. Mutation effect and efficiency were positively affected by mutagen number and dose concentration.

5.3 Sesame Mutant Library and Key Mutants

To our knowledge, a great number of new sesame mutants involved in seed yield, seed quality, plant growth and development rhythm, resistance (or tolerance) to abiotic and biotic stresses, or other key agronomic traits have been created via various mutagenesis techniques (Murty 1988; Kang et al. 1994; Wang et al. 2017; Zhang et al. 2019). In 1989, a narrow-leaf mutant was induced by the EMS from the M_4 progeny of var. Ven-52. The leaf vestige existed around the veins (Murty and Oropeza 1989). Pathirana (1991) identified 21 sesame lines induced by gamma ray induction and detected the high tolerance to phytophthora blight. 450 and 600 Gy could produce more lines with high tolerance to phytophthora blight than 100, 200, 300, and 750. Rajput et al. (1994) screened the population induced through γ-irradiation and found some mutants with early flowering date, short plant height, branching type, and high yield traits.

Notably, some mutants with determinate flowering habit, short internode, short stalk, short flowering period, male sterility, resistance to diseases and waterlogging, or other elite traits have been applied for genetics and breeding research in recent years (Cagirgan 2001; Abraham 2012; Zhang et al. 2016; Wang et al. 2017; Yu et al. 2017; Miao et al. 2020) (Table 5.1).

In the past decades, Chinese scientists constructed a sesame mutant library containing over 60,000 lines using a sibling of var. Yuzhi 11 through EMS mutagenesis approach (Wang et al. 2017; Zhang et al. 2019; Miao et al. 2020). Of these more than 1000 visible mutants involving dozens of agronomic and botanic traits have been identified and evaluated (Unpublished data, Haiyang Zhang). For the mutant library, 27 agronomic traits of the six trait groups including 'leaf type', 'plant type', 'flower and fertility', 'capsule and seed characters', 'other physical traits', and 'disease resistance and herbicide resistance' are applied to screen mutants. In particular, the determinate mutant DS899 (*dt*1 type) and the short-internode mutant Dw607 created by EMS exhibited the importance of mutagenesis technique in sesame breeding (Zhang et al. 2016; Miao et al. 2020). In 2015, based on the above two mutants, the first generation of Chinese sesame varieties with high adaption for mechanized cultivation, such as Yuzhi DS899 and Yuzhi Dw607 was bred by HAAS, China (Table 5.1).

5.4 Mutants Related to Botanic Traits

Previous reports described some sesame mutants involved in mutated leaf type, flower (corolla) morphology, flowering habit, flowering stage, and branching type (Murty and Oropeza 1989; Murty and Bhatia 1990; Kang et al. 1994; Boureima and Damme 2012; Madhusudan et al. 2013). For example, Madhusudan et al. (2013) found a mutant with split corolla from the M_2 generation of DS-1 through gamma radiation approach (500 Grey). Rajput et al. (1994) isolated some mutants with short structure, early flowering, uniculm, heavy branching, or high yield traits through γ-ray mutagenesis. Meanwhile, Boureima and Damme (2012) also performed γ-ray mutagenesis and detected a wide range of genetic variability with viable sesame mutants. In mutants, the main mutation included capsule indehiscence, petal broken, male sterility,

Table 5.1 Mutants and varieties induced by mutagenesis techniques in sesame

Mutation technique	Representative new sesame varieties or lines
Gamma radiation	ANK-S2, Ansanggae, Babil, Binatil-1, Birkan, Cairo White8, Eshtar, Rafiden, Sinai White48, Suwon 155, and Yuzhi 11
Chemical mutation	Kalika (BM3-7), Seodunkkae, UMA, USHA, Yangbaeckkae, Yuzhi 28, Yuzhi 29, Yuzhi 30, Yuzhi 31,Yuzhi DS899, Yuzhi Dw607, and Yuzhi Dw609
Space mutation	Jihangzhi 1, Jihangzhi 2, Jihangzhi 3, Jihangzhi 4, Zhengtaizhi 1, Zhongzhi 11, Zhongzhi 13
Recombination breeding and physical mutation	Pungsankkae, Suwonkkae, Zhengzhi 97C01, and Zhengzhi 98N09

Modified from Yu et al. (2017)

and seed coat color change, and yield variation. Especially, the 4 main yield-related traits, i.e., capsule size, capsule number per node, carpel number per capsule, and flowering date were improved in some mutants.

As to the flowering habit trait, changing the inflorescence indeterminacy to determinacy is one of the most important innovations for sesame breeding (Ashri 1998; Zhang et al. 2016). Differing from the normal varieties, there are two types of determinate mutants in sesame. One is *dt*2-typemutant induced through gamma radiation (5000 Grey) by an Israeli scientist Ashri A. (Ashri 1998) (Fig. 5.1a). The mutant only forms 3–5 capsule nodes per stem under different environments. The other is *dt*1-type determinate mutant (China Patent No. ZL2015108760163) created from a sibling line of Yuzhi 11 (used for the Sesame Genome Project) by EMS mutagenesis technique (Fig. 5.1b). The mutant can develop 8–20 capsule nodes. The capsule node number per stem varies with the altitude and sowing season. Trait comparison indicated that the two determinate mutants can form a terminal flower cluster on the top, as the shoot apical meristem was converted from continuous differentiation to the terminal meristem during maturing stage (Melzer et al. 2008; Zhang et al. 2016, 2019). Genetic analysis results revealed that the morphological difference between *dt*1 and *dt*2 mutants resulted from the allelic function of the *SiDt* gene (Zhang et al. 2016) (see more description in following chapters). At present, a series of determinate sesame varieties were bred with dt1 mutant as a parent using pedigree selection in China (Fig. 5.1c, d). In farms, all the determinate varieties derived from dt1 mutant present the consistency in capsule mature, the high tolerance to logging, and the high adaption for mechanization cultivation and harvest.

Plant height trait is also an important topic in sesame breeding. In 2009, a dwarf mutant Dw607 with short internode length was named induced from the sibling of var. Yuzhi 11 using EMS mutagenesis by Chinese sesame scientists (Fig. 5.2). Compared with the wild type, the internode length of the mutant decreased from 6.0–8.4 cm to 3.5–4.0 cm, which resulted in the significant reduction of the plant height (Miao et al. 2020). Interestingly, gene cloning results indicated that the target mutated gene *Sidwf1* (NCBI accession no. KY649623) regulating the dwarf and short internode length traits in sesame is homologous to the 'green evolution' gene in wheat (Miao et al. 2020) (see more description in following chapters). As to the first dwarf sesame variety Yuzhi Dw607, both the height to the first capsule and the plant height reduced, which result in the high tolerance to logging. As a result, the high capability of population density and capsule number exhibits the high yield potential in sesame production.

5.5 Mutants Related to Yield and Yield Components

In recent years, some mutants with the improvement of yield and yield components traits were obtained through γ-ray irradiation

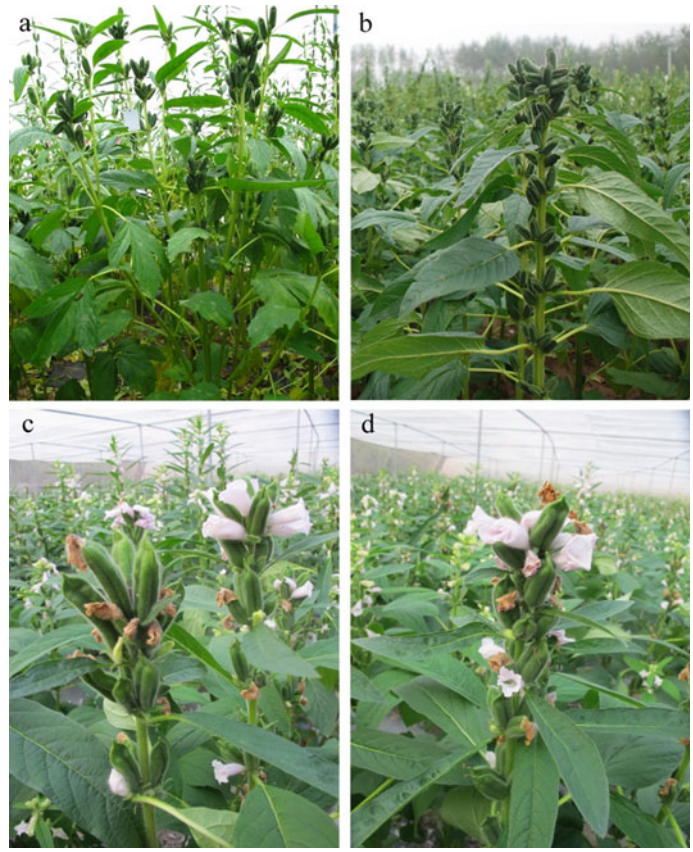


Fig. 5.1 Determinate flowering habit in sesame. **a**: *dt*2 type determinate mutant (reservation no. 08TP092). The *dt2* type mutant is induced by γ-rays (Ashri 1998). **b**: *dt*1 type determinate mutant DS899. **c**: Determinate variety Yuzhi 28 with determinate inflorescence and long capsule trait. **d**: Determinate variety Yuzhi 29 with determinate inflorescence and purple flower trait (Provided by Hongmei Miao)

Fig. 5.2 Dwarf mutant with short internode length in sesame. **a**: Comparison of the internode length between wild type (var. Yuzhi 11, left) and dwarf mutant (Dw607, right); **b**: Dwarf mutant Dw607 with short internode length and high capsule density; **c**: Dwarf variety Yuzhi Dw607 at maturing stage in Xinjiang experimental station, China (Provided by Haiyang Zhang)

(Murthy 1980; Al-Janabi et al. 1991; Tepora and Dela Cruz 1991; Potan et al. 1994; Rahman and Das 1994). Murthy (1980) screened 20 visible mutants in the M_2 generation exposed by γ-rays.

The results showed that 36–10, S337, and R149 presented significantly high yield. Meanwhile, 1-8, 2-2, 22-2, 34-3, and 43 maintained high oil content. Al-Janabi et al. (1991) selected the 11 promising mutants from the M_6 generation. The seed number per capsule, capsule number per plant, capsule number per leaf axil, and capsule length increased, while the branch number reduced in the mutants.

5.6 Mutants Related to Seed Quality Traits

Mutagenesis approaches have been applied for sesame seed quality breeding since the late twentieth century (Ashri 1994; Kang 1996; Begum and Dasgupta 2011). Ashri (1994) reported some sesame mutants with unique plant architecture, large seed size, and high seed quality induced by gamma radiation. Kang (1996) developed the mutant 'Suwonkkae' from the across derived from a local variety and a mutant induced by X-rays. The new line presented the superior quality than the parents. Moreover, Begum and Dasgupta (2011) reported some mutants with high oil content and high yield from the M_4 generation induced by γ-rays and EMS, respectively.

5.7 Mutants Related to Disease Resistance and Stress Tolerance

Shahin (1998) studied the effect of mutagens on the growth of *Aspergillus flavus* and the aflatoxin production in sesame seeds. The results showed that gammar irradiation could decrease the production of aflatoxin in the seeds. The irradiation of 2.5 kGy gamma rays could completely inhibited the production of aflatoxin in the inoculated sesame seeds under optimal conditions. Meanwhile, Cagirgan (2001) observed the tolerance to wilt disease in the M_2 mutations of the 4 sesame cultivars after irradiated with 150–750 Gy doses of gamma rays. In order to obtain the mutants with high tolerance to drought stress, Boureima et al. (2012) investigated the mutants induced by gamma radiation under polyethylene glycol (PEG) 6000 treatment during seedling stage and water deficiency treatment during vegetative stage, respectively. However, to our knowledge, no elite mutants with high resistance to stresses have been successfully applied for stress breeding in sesame.

5.8 Mutants Related to Male Sterility

Pathirana (1991) primarily reported a mutant with male sterility induced by gamma rays. In attempt to create male-sterile materials for hybridization breeding (Li and Chen 1998), irradiated dry seeds of sesame variety Yuzhi 4 with gamma rays (300, 500, and 700 Gy) of ^{60}Co. Ten mutants with male sterility mutation were obtained from the 60,000 M_2 individuals. Of which 6 mutants exhibited the characters of pure male sterility and female fertility. The male sterility could be inherited and would be controlled by single recessive gene. Moreover, Boranayaka et al. (2011) studied the pollen fertility of two genotypes after induced by gamma rays and EMS. The results indicated that either large or cryptic structural variation of pollens was observed in the mutants. The percentage of the fertile pollens in M_1 plants presented the reduced tendency with the increase of mutagen dosage.

References

Abraham B (2012) Quantitative trait variations through chemical mutagenesis in sesame (*Sesamum indicum* L.). M.Sc. Thesis, Haramaya University, Ethiopia

Al-Janabi KK, Wohayeb B, Khudhair H (1991) Mutation breeding in sesame (*Sesamum indicum* L.). In: IAEA, Proceedings of an International Symposium on the Contribution of Plant Mutation Breeding to Crop Improvement. Vienna, Austria, 18–22 June 1990, pp 393–398

Anbarasan K, Rajendranmkk R, Sivalingam D (2015) Studies on the mutagenic effectiveness and efficiency of gamma rays, EMS and combined treatment in

sesame (*Sesamum indicum* L.) var.TMV3. Res J Pharmaceut Biol Chem Sci 6(4):589–595

Anbarasan K, Rajendranmkk R, Sivalingam D, Chidambaram ALC (2014) Studies on the effect of EMS and colchicine in M1 generation of sesame (*Sesamum indicum* L.) Var. TMV3. Int lett Nat l Sci 11(2):209–214

Anbarasan K, Sivalingam D, Rajendran R, Anbazhagan M, Chidambaram AA (2013) Studies on the mutagenic effect of EMS on seed germination and seedling characters of sesame (*Sesamum indicum* L.) Var. TMV3. Int J Res Biol Sci 3(1):68–70

Ashri A (1994) Modification and adaptation of the induced determinate sesame mutant by cross-breeding and its evaluation. In: Mutation Breeding of Oil Seed Crops. Proc FinalFAO/IAEA Research Co-ordination Meeting, 11–15 Jan. 1993, IAEA, Vienna, Austria, IAEA-TECDOC-781, pp 111–114

Ashri A (1998) Sesame breeding. In: Janick J (ed) Plant Breeding Reviews, vol 16. Wiley, Israel, pp 179–228

Ashri A (2001) Induced mutations in sesame breeding. No. IAEA-TECDOC-1195

Begum T, Dasgupta T (2010) A comparison of the effects of physical and chemical mutagens in sesame (*Sesamum indicum* L.). Genet Mol Biol 33(4):761–766

Begum T, Dasgupta T (2011) Effect of mutagens on character association in sesame (*Sesamum indicum* L.). Pak J Bot 43(1):243–251

Begum T, Dasgupta T (2015) Ameloration of seed yield, oil content and oil quality through induced mutagenesis in sesame (*Sesamum indicum* L.). Bangladesh J Bot 44(1): 15–22

Boranayaka MB, Gowda RK, Nandini B, Satish RG, Pujer SB (2010) Influence of gamma rays and ethyl methanesulphonate on germination and seedling survival in sesame (*Sesamum indicu*m L.). Int J Plant Sci 5:655–659

Boranayaka MB, Ibrahim SM, Rajesha G, Shadakshari TV, Ashoka KR et al. (2011) Gamma rays and ethylmethane sulphonate induced pollen sterility in sesame (*Sesamum indicum* L.). Int J Plant Sci 6 (2):240–242

Boureima S, Oukarroum A, Diouf M, Cisse N, Damme PV (2012) Screening for drought tolerance in mutant germplasm of sesame (*Sesamum indicum*) probing by chlorophyll a fluorescence. Environ Exp Bot 81(3):37–43

Boureima S, Damme SPV (2012) Sesame (*Sesamum indicum* L.) improvement by induced mutation: effect of mutagenesis on drought tolerance and productivity. Afrika Focus 25(1):73–102

Cagirgan MI (2001) Mutation techniques in sesame (*Sesamum indicum* L.) for intensive management: confirmed mutants. In: Van Zanten L (ed) Sesame Improvement by Induced Mutations, Final FAO/IAEA Co-ord. Res. Mtng, Vienna, Austria, pp 31–40

Diouf M, Boureima S, Diop T, Airgan MI (2010) Gamma rays-induced mutant spectrum and frequency in sesame. Turk J Field Crops 15(1):99–105

Gong HH, Zhao FT, Pei W, Meng QH (2016) Advances in sesame (*Seamum indicum* L.) germlasm resources and molecular biology research. J Plant Genet Resour 17(3):517–522

Kang CW (1996) Breeding sesame disease shatter resistance and high yielding varieties through induced mutations. Proc Second FAO/IAEA Research Co-ord Meet, On induced mutations for sesame improvement, 9–3 Sept. 1996, Antalya, Turkey

Kang CW, Lee JI, Choi BH (1994) Mutation breeding for disease resistance and high yield of sesame (*Sesamum indicum* L.) In: Mutation Breeding of Oil Seed Crops, Proc Final FAO/IAEA Research Co-ordination meeting, 11–15 Jan. 1993, IAEA, Vienna, Austria, IAEA-TECDOC-781, pp 69–82

Kumar G, Yadav RS (2010) EMS induced genomic disorders in sesame (*Sesamum Indicum* L.). Rom J Biol-Plant Biol 55(2):97–104

Kumar P, Madhusudan K, Nadaf HL, Patil RK, Deshpande SK (2012) Combining ability and gene action studies in inter-mutant hybrids of sesame (*Sesamum indicum* L.). Karnataka J Agric Sci 25(1):1–4

Kumari V, Chaudhary HK, Prasad R, Kumar A, Singh A, Jambhulkar S, Sanju S (2016) Frequency and spectrum of mutations induced by Gamma radiations and Ethy1 methane sulphonate in sesame (*Sesamum indicum* L.). Sci Agri 14(3):270–278

Langham DR (2007) Phenology of sesame. In: Janick J, Whipkey A (eds) Issues in New Crops and New Uses. ASHS Press, Alexandria, VA, pp 144–182

Li YD, Chen QM (1998) Induction of male sterility mutation in sesame (*Sesamum indicum* L.) and a preliminary study of the genetics of the mutants. Chine J Oil Crop Sci 20(1):24–27

Lv HX, Zhang YX, Wang LH, Wei WL, Li DH et al (2011) Sesame mutant induced by space flight treatment and analysis of polymorphism by AFLP molecular marker. J Nulear Agri Sci 25(5):851–856

Madhusudan K, Nadaf HL, Motagi BN (2013) Split corolla mutant–an useful genetic stock bred through induced mutagenesis in Sesame (*Sesamum indicum* L.). Karnataka J Agric Sci 26(3):417–418

Melzer S, Lens F, Gennen J, Vanneste S, Rohde A et al (2008) Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana*. Nat Genet 40:1489–1492

Miao HM, Li C, Duan YH, Wei LB, Ju M, Zhang H (2020) Identification of a *Sidwf1* gene controlling short internode length trait in the sesame dwarf mutant *dw607*. Theor Appl Genet 133:73–86

Muhammad LM, Falusi OA, Daudu OAY, Gado AA, Lateef AA et al. (2013) Radiation induced polygenic mutation in two common Nigerian sesame (*Sesamum*

indicum L.) cultivars. Int J Biotechnol Food Sci 1 (2):23–28

Murthy GSS (1980) Promising sesame mutants. In: Mutation Breeding Newsletter, 1980

Murty GSS (1988) Inheritance of three new mutants in sesame. Curr Sci 57(4):204–206

Murty GSS, Bhatia CR (1990) Inheritance of polypetalous corolla mutation in sesame. Proc Indian Acad Sci (Plant Sci) 100(1):7–10

Murty BR, Oropeza F (1989) An induced leaf differentiation mutant in *Sesamum indicum* L. Curr Sci 58 (8):464–466

Pathirana R (1991) Induced mutations in sesame facilitate adapting crop to intensive agriculture. In: EUCARPIA Symposium on Genetic Manipulation in Plant Breeding, Tarragona, Spain, May 26–30

Pathirana R (1994) Induced mutations and anther culture for sesame improvement. In: Proceedings of a final Research Co-ordination Meeting of an FAO/IAEA Co-ordinated Research Programme, Vienna, Austria, January, 11–15

Pei XY, Li CM, Qin LL (2014) Sesame research status and development trend analysis on basis of foreign periodical literature in English. J Henan Agric Sci 43 (5):13–19

Potan N, Rungsipiyakul S, Dao-Ngarm T, Chareonart S (1994) Improvement of sesame by induced mutation in Thailand. Mutation Breeding of Oil Seed Crops, Proc Final FAO/IAEA Research Co-ordination Meeting, 11–15 Jan. IAEA, Vienna, Austria, pp 83–88

Qiu CP, Zhang HY, Chang SX, Wei LB, Miao HM (2014) Laboratory detecting method for pathogenicity of *Fusarium oxysporum* Schl. f. sp. sesame isolates. Acta Phytopathol Sin 44(1):26–35

Rahman A, Das ML (1994) Evolution of improved varieties of sesame through induced mutation. In: Mutation Breeding of Oil Seed Crops. Proc. Final FAO/IAEA Research Co-ordination Meeting, 11–15 Jan 1993, IAEA, Vienna, Austria, IAEATECDOC-781, pp 115–118

Rajput MA, Sarwar G, Siddiqui KA, Siddiqui MA (1994) Genetic improvement of sesame for plant architecture and grain yield through nuclear techniques. In: Mutation Breeding of Oil Seed Crops, Proc Final FAO/IAEA Research Co-ordination Meeting, 11–15 Jan 1993, IAEA, Vienna, Austria, IAEATECDOC-781, pp 89–96

Ravichandran V, Jayakumar S (2015) Mutagenic effectiveness and efficiency of gamma rays and EMS in sesame (*Sesamum indicum* L.). Int J Res Bot 5(2):14–19

Ramadoss BR, Ganesamurthy K, Angappan K, Gunasekaran M (2014) Mutagenic effectiveness and efficiency of gamma rays in sesame (*Sesamum indicum* L.). Glob J Mol Sci 9(1):1–6

Salisu N, Adamu AK, Mu'Azu S, Dangora DB (2011) Effect of colchicine induced mutagenesis on growth and yield of sesame (*Sesamum indicum* L.). Bayero J Pure Appl Sci 4(1):121–125

Shahin AAM (1998) Effect of different factors including gamma radiation on the growth and aflatoxin production of *Aspergillus flavus* in sesame seeds. Egyp J Microbiol 33(3):455–468

Shi S (1991) Application of mutagenesis technique in the improvement of sesame varieties. Chin J Oil Crop Sci 2:93–96

Tepora NM, Dela Cruz QD (1991) Mutation induction in sesame (*Sesamum indicum* L.). In: IAEA, Proceedings of an International Symposium on the Contribution of Plant Mutation Breeding to Crop Improvement, Vienna, Austria, 18–22 June 1990, pp 371–374

Uzun B, Çağırgan Mİ (2009) Identification of molecular markers linked to determinate growth habit in sesame. Euphytica 166:379–384

Van Zanten L (2001) Sesame improvement by induced mutations: Results of the co-ordinated research project and recommendation for future studies. In: Van Zanten L (ed) Sesame Improvement by Induced Mutations 1. Proc Final FAO/IAEA Co-ord. Res. Mtng, IAEA, Vienna, Austria, pp 1–12

Wang H, Zhang H, Ma Q, Wei L, Ju M et al (2017) Optimization of EMS mutagenesis condition and screening of mutants in sesame. J Henan Agric Sci 46(1):36–41

Yu M, Zhou QF, Zhao JG, Zhang GG (2017) Research advances on induced mutation breeding of sesame. Biotechnol Bull 33(11):8–12

Zhang HY, Miao HM, Li C, Wei L, Duan YH et al (2016) Ultra-dense SNP genetic map construction and identification of *SiDt* gene controlling the determinate growth habit in *Sesamum indicum* L. Sci Rep 6:31556

Zhang H, Miao H, Ju M (2019) Potential for adaptation to climate change through genomic breeding in sesame. In: Kole C (ed) Genomic Designing of Climate-Smart Oilseed Crops. Springer, Cham, Switzerland, pp 374–376

Tissue Culture and Genetic Transformation in Sesame

6

Hongmei Miao, Ming Ju, Huili Wang, and Haiyang Zhang

Abstract

Plant tissue culture and genetic transformation are powerful biotechniques for plant genetic improvement. For sesame, in vitro callus induction and differentiation with high efficiency are very difficult because of the specific characters of the species, and employing genetic transformation is too difficult. In the past forty years, many scientists tried a huge number of experiments in order to establish efficient callus induction and plantlet regeneration system in sesame. In this section, we present the history and the main progresses of sesame tissue culture and gene transformation research. Several successful examples for plant regeneration and gene transformation in sesame are listed. In addition, the main factors influencing the efficiency of tissue culture, plantlet regeneration, and gene transformation are discussed. The application prospect of tissue culture and gene transformation in sesame genetics and breeding research is also introduced.

6.1 Introduction

Sesame belongs to the Pedaliaceae family and is originated from the tropical regions. Its rapid growth and development and short life-cycle (low to 70 d) character exhibit high potential for sesame to be a model crop for genetics research in key agronomic traits. However, sesame is regarded as one of the most recalcitrant crops for in vitro callus induction and differentiation (Were et al. 2006; Yadav et al. 2010; Raja and Jayabalan 2011; Zhang et al. 2019). In the past forty years, many researchers conducted technical optimization experiments in order to establish the effective techniques of tissue culture and plantlet regeneration in sesame. Early studies involved adventitious shoot induction techniques (George et al. 1987; Chen et al. 1994; Zhi et al. 1998; Saravanan and Nadarajan 2005; Seo et al. 2007; Chattopadhyaya et al. 2010; An 2009; An et al. 2011). Later, in order to increase the efficiency of multiple shoot formation and plantlet regeneration, many factors including genotype, explant type, explant pre-treatment method, culture medium and condition, and plant hormones and the concentration combinations have been investigated (Shi and Cai 1986, 1989a; Yi et al. 1997; Were et al. 2006; An 2009; Miao et al. 2012). At the same time, both *Agrobacterium*-mediated transformation and particle bombarding methods were also applied in sesame in order to set up the gene transformation technique (Chen

H. Miao · M. Ju · H. Wang · H. Zhang (✉)
Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, People's Republic of China
e-mail: zhanghaiyang@zzu.edu.cn

H. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-98098-0_6

et al. 1996; Yadav et al. 2010; Chowdhury et al. 2014). In recent a few years, the techniques of callus induction, differentiation, and plantlet regeneration have been established in sesame and are being applied for sesame transgene research (Mary and Jayabalan 1997; Were et al. 2006; Charaborti and Ghosh 2009). Fortunately, the *Agrobacterium*-mediated gene transformation system has been successfully established in sesame in recent a few years (Mitsuma et al. 2004; Chowdhury et al. 2017; Zhang et al. 2019), even though the transformation efficiency is still low. The findings provide the powerful technique system for exploring the molecular genetic basis of important agronomic traits and expanding the genetic basis of the cultivated sesame in near future.

6.2 Tissue Culture and Plant Regeneration in Sesame

For sesame, studies of tissue culture and plant regeneration techniques initiated in the 1980s (Liu 1986; Liu et al. 1994; Qu et al. 1994; Li et al. 1996; Mary and Jayabalan 1997; Xu et al. 1997; Zhi et al. 1998; Baskaran and Jayabalan 2006; Wei et al 2007; Charaborti and Ghosh 2009; Cui et al 2010). At present, the tissue culture and plant regeneration techniques in sesame are mainly applied for two different purposes. One is to produce adventitious shoots and propagate valuable plants via explant induction, which is designated as organogenesis (direct way). The other is to form adventitious shoots or embryos via callus tissue induction, differentiate the callus tissues, and then develop regenerated plants, which is designated as somatic embryogenesis (indirect way) (Miao et al. 2012).

In the early stage of the tissue culture technology development in sesame, many scientists focused on the adventitious shoot induction techniques (George et al. 1987; Chen et al. 1994; Zhi et al. 1998; Saravanan and Nadarajan 2005; Seo et al. 2007; Chattopadhyaya et al. 2010; An 2009; An et al. 2011). Several reports showed that various explants such as seed, shoot tip, stem, and young leaf could be used to directly induce plexus bud and for propagation (George et al. 1987, 1989; Kwon et al. 1993; Rao and Vaidyanath 1997; Gangopadhyay et al. 1998; Were et al. 2006; Lokesha et al. 2007; Seo et al. 2007; Ahmed et al. 2008; Abdellatef et al. 2010; Chattopadhyaya et al. 2010; Raja and Jayabalan 2010; Honnale 2011; Raja and Jayabalan 2011; Shashidhara et al. 2011; Singh and Shagufta 2011; Lokesha et al. 2012; Wadeyar 2013; Malaghan et al. 2016; Pratik et al. 2016). However, inducing shoots from the above explants via callus induction and differentiation in sesame seems very difficult (Taskin and Turgut 1997). Till now, most of the regenerated plants in some sesame genotypes were induced from the adventitious buds which are induced through callus tissue induction and differentiation methods (Shi and Cai 1986, 1989a; Li et al. 1996; Miao et al. 2012; Bangaremma et al. 2013).

6.2.1 Somatic Embryo Formation

Somatic embryogenesis belongs to indirect way to realize the plant regeneration. As to the various explants, switching organ growth to the dedifferentiation and callus formation is the first culture strategy. Even though somatic embryos could form from callus tissues induced from hypocotyl explants, the differentiation of somatic embryos into new sesame plantlets is difficult (Mary and Jayabalan 1997). The failure or low rate of callus differentiation, low repeatability of successful methods, or significant effect of genotypes could result into the low ratio of embryo formation in sesame (Mary and Jayabalan 1997; Were et al. 2006; Charaborti and Ghosh 2009; Tiwari et al. 2011). Miao et al. (2012) investigated the effects of explant type, hormone type, and hormone concentration on callus induction and plantlet regeneration from various genotypes of the cultivated sesame (*S. indicum*), *S. radiatum*, *S. schinzianum*, and hybrids of the interspecific cross *S. schinzianum* × *S. indicum*, and established a highly efficient system of callus induction, differentiation, and plantlet regeneration for the *Sesamum*

genus. Meanwhile, George et al. (1987) detected the embryo-like structure of callus tissues which were derived from hypocotyl segments. However, the embryo-like cells did not form shoots. Xu et al. (1997) obtained somatic embryos on the surface of the induced callus tissues. The somatic embryos in late maturation stage formed cotyledon, shoot apex, and radical tissue after several rounds of subculture. Liu et al. (1994) successfully induced somatic embryos from embryonic callus tissues through suspension shock culture. They also investigated the variation of hormone level during the somatic embryo development. The results showed that the endogenous hormone content in cells increased, as the embryonic callus formed, while the exogenous hormones are necessary to induce explant de-differentiation. In addition, main research results suggested that the genotype was the key factor to limit the application of callus induction and plantlet regeneration methods in sesame (Mary and Jayabalan 1997; Were et al. 2006; Charaborti and Ghosh 2009).

6.2.2 Genotype

Genotype is a key factor affecting the efficiency of tissue culture techniques in sesame. In the previous reports, many sesame germplasm accessions and wild *Sesamum* species including *S. schinzianum*, *S. radiatum*, and *S. occidentalis* have been studied for plant regeneration (Table 6.1). Different genotypes exhibited different capacity of plant regeneration. Kwon et al. (1993) compared the callus induction rate of two cultivars and *S. orientate* which was collected from Ghana. The results showed that *S. orientate* exhibited significantly higher rate than cultivated sesame. Rao and Vaidyanath (1997) compared the nine sesame varieties with the wild species *S. occidentalis* cultured in the studied medium. The results indicated that the average shoot number of *S. occidentalis* was much higher. Miao et al. (2012) examined the genotype effect of three cultivars, two wild species *S. radiatum* and *S. schinzianum*, and five distant hybrid lines of *S. schinzianum* × *S. indicum* on callus induction and shoot regeneration. The results confirmed that genotype determined the callus induction rate, callus quality, and the final plant regeneration rate. The highest frequency of callus inducement and differentiation in the wild species was 97.50% and 94.02%, respectively, higher than those in cultivars (highest to 40.60% and 8.16%, respectively). Interestingly, the distant hybrids showed medium frequency of callus inducement (46.67%) and differentiation (89.29%) because of their hybridized genetic background. Therefore, optimizing the regeneration technique for different genotypes is necessary in sesame tissue culture research. George et al. (1989) performed the shoot induction using seven cultivars. Of the seven cultivars, the maximum number of the produced multiple buds varied from 15 to 18 per explants, and 80% explants could develop buds. The minimum of the produced multiple buds was 4–5, and only 44% explants developed buds. The genotype of donors indeed could influence in vitro propagation of explants, although shoot tip induction is a reliable method in sesame.

6.2.3 Explant Type

In sesame, many types of explants such as immature embryo, mature embryo, anther, seed, hypocotyl, cotyledon, shoot tip, stem segment with axillary, de-embryonated cotyledon, root, and transverse thin cell layer of the internodes have been tested to induce callus or adventitious shoot (Table 6.1). The induction frequency of callus tissues and plant regeneration varied with the various explant types and sources. In general, the induction rate of callus tissues of hypocotyls is significantly higher than cotyledon and other explants. Meanwhile, the efficiency of callus induction and regeneration was affected by other factors, including explant age and collection time, development stage of donor plants, nutritional condition, and cultivation condition. Liu et al. (1990) found that anthers of plants cultured in autumn were more suitable for callus induction than summer planted materials. To rescue the immature embryos, especially from the

Table 6.1 An overview of the tissue culture and plant regeneration techniques in sesame

Genotype	Explant type	Culture results	References
S. indicum	Hypocotyl and shoot tip with cotyledon	Hypocotyl-embryo-like tissue produced from callus tissue; shoot tip with multiple cotyledon shoots	George et al. (1987)
S. indicum	Hypocotyl	Bud differentiated from callus tissue	Shi and Cai (1989a)
S. indicum	Shoot tip	Multiple shoot buds	George et al. (1989)
S. indicum	Anther	Bud differentiated from callus tissue	Shi and Cai (1989b)
S. indicum	Immature embryo	Single plantlet	Sudhaker et al. (1990)
S. indicum	Anther	Callus tissue and differentiated root from callus tissue	Liu (1990)
S. orientale (*S. indicum*)	Hypocotyl and cotyledon	Shoot differentiated from callus tissue	Li and Chen (1990)
S. indicum	Hypocotyl and cotyledon	Embryo-like structure formed in hypocotyl-derived calli	Kwon et al. (1993)
S. orientale (*S. indicum*)	Hypocotyl and cotyledon	Adventitious shoot from hypocotyl- and cotyledon-derived calli; multiple shoots	
S. indicum	Seed	Multiple shoots	Chen et al. (1994)
S. indicum	Hypocotyl	Embryo-like structure	Liu et al. (1994)
S. indicum × *S. Schinziamum*	Immature embryo	Shooting	Qu et al. (1994)
S. indicum	Hypocotyl	Callus tissue; multiple shoots	Li et al. (1996)
S. indicum; *S. occidentalis*	Seedling shoot tip	Multiple shoots	Rao and Vaidyanath (1997)
S. indicum	Hypocotyl	Somatic embryos induced from hypocotyl-derived callus	Mary and Jayabalan (1997)
S. indicum	Cotyledon, root, and subapical hypocotyl segment	Embryos present on the callus surface; somatic embryos with cotyledon, shoot apex, and radicle tissue	Xu et al. (1997)
S. indicum	Hypocotyl, cotyledon, and radicle	Hypocotyl budding differentiated from callus tissue and embryonic callus with differentiation potential; cotyledon shoot; radicle callus tissue	Yi et al. (1997)
S. indicum	Node	Multiple shoot	Gangopadhyay et al. (1998)
S. indicum	Hypocotyl, cotyledon, and cotyledon node	Callus tissue	Yu (2005)
S. indicum	De-embryonated cotyledon	Adventitious shoot	Seo et al. (2007)
S. indicum	Leaf	Shoot	Raja and Jayabalan (2010)
)S. indicum	Shoot tips	Adventitious shoot	Abdellatef et al. (2010)
S. indicum		Shoot bud	

(continued)

Table 6.1 (continued)

Genotype	Explant type	Culture results	References
	Transverse thin cell layer of the internodes		Chattopadhyaya et al. (2010)
S. indicum	Hypocotyl and cotyledon	Hypocotyl callus; cotyledon-direct regeneration	Honnale (2011)
S. indicum	Hypocotyl	Shoot regenerated from callus	Wadeyar and Lokesha (2011)
S. indicum	De-embryonated cotyledon	Adventitious shoot	Lokesha et al. (2012)
S. indicum	Adventitious shoot	De-embryonated cotyledonary explant	Malaghan et al. (2013)
S. indicum	Anther	Shoot	Yifter et al. (2013)
S. indicum	De-embryonated cotyledon	Shoot	Pratik et al. (2016)
S. indicum	De-embryonated cotyledon	Shoot	Malaghan et al. (2016)
S. radiatum	Stem with axil	Seedling regenerated from callus tissue	Shi and Cai (1986)
S. indicum	Hypocotyl and cotyledon	Callus tissue; multiple shoots	Miao et al. (2012)
S. radiatum			
S. schinziamum			
S. indicum × *S. radiatum*			

interspecific hybrid seeds, about 10-day-old embryos after pollination are ideal for plant regeneration (Unpublished data, Haiyang Zhang).

6.2.4 Plant Growth Regulators

Besides the above factors, plant growth regulators and nutrition substances are also the main factors determining the tissue culture efficiency in sesame. George et al. (1987) cultured the hypocotyls and cotyledon explants on MS medium without hormones and found all the explants necrosed without any morphogenetic response. In sesame, common cytokinins, such as 6-BA (6-benzylaminopurine), KT (kinetin), TDZ (thidiazuron, B_1), and ZT (zeatin) and auxins, such as 2,4-D (2,4-dichlorophenoxy acetic acid), NAA (naphthalene acetic acid), IAA (indole acetic acid), and IBA (indole-3-butyric acid) are widely applied for tissue culture (George et al. 1987; Were et al. 2006; Ahmed et al. 2008; Charaborti and Ghosh 2009; Raja and Jayabalan 2011). In addition, some growth regulating substances, such as GA_3 (gibberellic acid), $AgNO_3$, and activated charcoal are also added in culture media to regulate the growth and development of various tissues. Studies showed that the specific hormone combinations with various concentrations could induce different morphogenetic responses of explants in sesame (Baskaran and Jayabalan 2006). Singh et al. (2006) found that both 2,4-D and 6-BA could reduce the formation of callus tissue in sesame. Miao et al. (2012) also detected that 2,4-D had the function in promoting

the cell division and expansion which took the important function in increasing the callus induction efficiency. However, the induced calli were apt to forming roots, rather than shoot differentiation. Meanwhile, 6-BA could accelerate cell division and were favorable for the formation of high-quality calli in sesame.

As to hormone combination, George et al. (1987) studied the function of cytokinin and auxin combinations in sesame. When NAA (0.2 $mg{\cdot}L^{-1}$) in combination with BA, KIN, ZEA, or 2iP was used, the callus tissues with moderate growth rate were observed at the fourth week after culture. George et al. (1989) found that shoot tip with cotyledon pretreated with cytokinins showed high induction frequency of multiple buds, of which 6-BA (8 $mg{\cdot}L^{-1}$) was the most favorable cytokinin to form multiple buds, while 2ip (8 $mg{\cdot}L^{-1}$) exhibited the lowest effect.

Meanwhile, Kwon et al. (1993) screened a combination of 1–2 $mg{\cdot}L^{-1}$ NAA and 0.2–0.6 $mg{\cdot}L^{-1}$ BAP which was effective for callus induction from hypocotyl and cotyledon explants. The embryo-like structures were formed in hypocotyl-derived calli on the medium supplemented with only 1 $mg{\cdot}L^{-1}$ 2,4-D. Addition of casein hydrolysate (1–2 $g{\cdot}L^{-1}$) to the regeneration media containing 0.1 $mg{\cdot}L^{-1}$ NAA plus 1–4 $mg{\cdot}L^{-1}$ BAP could effectively increase the formation rate of adventitious shoots from the hypocotyl- (22–42%) and cotyledon-derived calli (16–34%), respectively, in *S. orientale*. In general, high concentration of BAP (3–4 $mg{\cdot}L^{-1}$) in combination with casein hydrolysate could favor the formation of multiple shoots, while low concentration (1–2 $mg{\cdot}L^{-1}$) induced the formation of single shoot. Under the condition of low BAP concentration, addition of casein hydrolysate with high concentration also tended to induce the formation of multiple shoots. Meanwhile, the differentiation of roots was easily induced from the regenerated shoots on 1/2 MS medium supplemented with 0.5 $mg{\cdot}L^{-1}$ NAA. Till now, there are rarely any reports on the mechanism of hormone function in sesame callus differentiation. To set up the stable and high-efficient callus induction and differentiation system in sesame, more studies of explant dedifferentiation regulation and the response mechanism of explants on hormone combinations should be conducted in future.

6.2.5 Culture Medium

In the process of in vitro tissue culture, various culture media with different nutritional components and concentrations should be supplied for specific tissues at specific growth stages. MS medium is the most common medium applied for sesame tissue culture, while 1/2MS sometimes was used for seed germination and root formation of the regenerated shoots (Yi et al. 1997; An et al. 2011; Miao et al. 2012). Other media such as B_5, Nitsch, LS, and White have been applied for sesame tissue culture and plant regeneration. Li and chen (1990) found that MS, B_5, and LS medium gave insignificant effects on the induction of sesame callus tissue. Yi et al. (1997) detected that the MS medium was the most optimum medium for callus induction of black sesame. Rao and Vaidyanath (1997) investigated the induction of multiple shoots in MS, LS, Camborg's, and L_6 medium, respectively. The induction percentage of multiple shoots varied from 38 to 100% for different genotypes on different media. Percentage of multiple shoot response of genotype T-85, TC-25, Co-I, and Rajeshwari reached 100% on all the four types of media. The lowest percentage of 38% was for the genotype Krishna on B_5 medium.

6.2.6 Other Additives and Treatments

In addition to the above main factors such as genotype, plant growth regulator, and explants which affect callus induction and shoot regeneration, some other additives and treatments such as glutamine, casein hydrolysate, L-tyrosine, activated charcoal, and ethylene inhibitors including $AgNO_3$ and $CoCl_2$ also have certain effects on sesame tissue culture. Shi and Cai (1986) added casein hydrolysate in the

regeneration induction medium for sesame stem section culture and found its benefit for shoot regeneration. Kwon et al. (1993) also found that addition of casein hydrolysate effectively increased the formation rate of adventitious shoots from hypocotyl- and cotyledon-derived calli in *S. orientale*. In sesame anther culture, addition of 300 mg·L^{-1} L-tyrosine did not promote the formation of pollen callus but presented toxic function for explant culture.

In addition, Abdellatef (2010) studied the effect of ethylene inhibitors $AgNO_3$ and $CoCl_2$ on sesame tissue culture and found that $AgNO_3$ increased the number of shoots from 2.7 to 3.7 per explant, and the shoot length increased from 1.3 to 2.9 cm accordingly. Meanwhile, addition of 5 mg·L^{-1} $CoCl_2$ induced the root length from 3.5 cm to 17 cm. The stimulative effect might result from a reduction of ethylene concentration or inhibition of ethylene action.

Xu et al. (1997) found that addition of 0.5% activated charcoal and 3% mannitol medium had good effects on embryos development. However, Shi and Cai (1989b) detected that addition of 0.2% activated carbon took no effects on callus induction and regeneration for sesame anther culture.

Meanwhile, some scientists found a few culture treatments could give aid to optimize the tissue culture techniques. To study the effect of cytokinins on multiple bud induction in the various sesame cultivars, George et al. (1989) soaked the seeds in the 6-BA, zeatin, or 2iP solution, respectively for 72 h and then put all the treated seeds on MS medium for germination and further multiple shoot induction. As a result, the induction frequency of the multiple shoot buds from the explants pretreated with cytokine increased. Rao and Vaidyanath (1997) considered that the total number of shoots induced from the explants pretreated by GA_3 was higher than those of control and the seedlings pretreated with BAP in different varieties and wild species. Moreover, dark treatment was also detected as a crucial step for shoot regeneration from cotyledon in sesame.

6.3 Application of Tissue Culture Techniques in Interspecific Hybridization in Sesame

Plant tissue culture technique is developed based on the theory of plant cell totipotency and the specific character of tissue regeneration. At present, the tissue culture techniques in sesame are mainly applied for rescuing the easily abortive embryos and precious germplasm materials, overcoming the incompatibility of distant hybridization in interspecific hybridization, and facilitating genetic modification. To improve resistance level of sesame to biotic and abiotic stresses, scientists performed interspecific hybridization research since 1950s (Joshi 1961). However, the low frequency of interspecific hybrids and the high incompatibility between the cultivated sesame and most of the wild *Sesamum* species limit the application of the elite genotypes of the wild species in sesame breeding (Joshi 1961; Subramanian 2003; Nimmakayala et al. 2011; Zhang et al. 2013, 2019; Yang et al. 2017). Only the progeny derived from the cross between *S. malabaricum* ($2n = 26$) and the cultivated sesame exhibited high fertility rate under natural growth conditions (Nimmakayala et al. 2011). Moreover, to obtain hybrids from the crosses between the cultivated sesame and other remote wild *Sesamum* species and interspecific hybrids with different chromosome karyotypes, such as *S. schinzianum* ($2n = 64$), *S. radiatum* ($2n = 64$), *S. alatum* ($2n = 26$), and *S. indicatum* (hybrid species derived from *S. indicum* and *S. prostratum*) (Mukherji 1947), the embryo rescue and tissue culture techniques should be applied (Nimmakayala et al. 2011; Miao et al. 2012; Zhang et al. 2013; Yang et al. 2017; Zhao et al. 2018). To assess the interspecific hybrids, some specific simple sequence repeat (SSR) markers have been screened and utilized (Chinese patent no. ZL. 201210318496.8). The findings thus give impetus to the interspecific hybridization breeding in sesame.

6.4 Genetic Modification in Sesame

Plant genetic transformation system includes the recombinant DNA technology and the tissue culture skills, which introduce the exogenous genes or foreign DNA fragments into the receptor genomes and transmits the new genetic information into the offsprings. Sesame is one of the most recalcitrant crops for tissue culture and plantlet regeneration. Thus, compared with cereals and other oilseed crops, the development of genetic transformation techniques in sesame is relatively lagging. In order to develop an efficient genetic modification technique in sesame, many researchers focused on exploring the genetic modification techniques, especially the *Agrobacterium*-mediated transformation and particle bombarding methods via adventitious shoot formation (Chen et al. 1996; Yadav et al. 2010). In the past three decades, the main transgenic methods such as biolistic or particle gun bombardment, *Agrobacterium*-mediated transformation, pollen-tube channel transformation, and explant dipping methods including floral dipping, pollen infiltration, and suspension drop have been used in sesame (Table 6.2). Of these the former two transformation methods depend on the tissue culture and shoot regeneration system.

6.4.1 Genetic Modification Via Micro-Particle Bombardment in Sesame

As a direct gene transfer method for forming transgenic plants, micro-particle bombardment technique was developed in the 1980s. Till now, the technique is widely applied for transgene delivery in various crops (Christou 1995; Breitler et al. 2002). Particle bombardment can directly transfer the foreign DNA molecules into the cells or tissues, independent of cell or tissue type. Chen et al. (1996) performed particle bombardment using cotyledon as explant with the aid of the direct multiple shoot induction techniques. Bhattacharyya et al. (2015) also analyzed the effect of particle bombardment transformation technique on 5-day-old apical tissues of sesame seedlings (cv. Rama). The results of PCR, RT-PCR, western blot, and enzymatic assay showed that four transformants carrying the exogenous *bar* (bialaphos resistance) gene with single integration site in the genome. The mean transformation frequency touched 15.84%.

6.4.2 *Agrobacterium*-Mediated Genetic Modification in Sesame

Differing from the particle bombardment method, *Agrobacterium*-mediated genetic transformation is easy to obtain more transformants with low copy of exogenous genes. Some advantages such as high conversion efficiency and low equipment cost make the *Agrobacterium*-mediated genetic transformation widely used in gene function research, biological metabolite production, and transgenic breeding. For example, the *Agrobacterium rhizogenes*-mediated transformation is mainly used to culture the transformed hairy roots as a plant organ bioreactor and to produce valuable substances such as the recombinant proteins and pharmaceutically important secondary metabolites (Jin et al. 2005).

In the past two decades, scientists applied the tissue induction and plantlet regeneration techniques in the *Agrobacterium*-mediated gene transformation in sesame (Taskin et al. 1999; Yadav et al. 2010; Chowdhury et al. 2014). Taskin et al. (1999) confirmed that sesame was an amenable crop for *Agrobacterium*-mediated transformation; however, no transformants generated due to low efficient regeneration system in sesame. Yadav et al. (2010) established the *Agrobacterium*-mediated transformation technique in sesame for the first time, based on the adventitious shoot induction technique. PCR, RT-PCR, and southern blot analysis showed that the transformation frequency reached up to 1.01%. Meanwhile, Chowdhury et al. (2014) performed the *Agrobacterium*-mediated transformation using de-embryonated cotyledons at

Table 6.2 An overview of the genetic transformation techniques in sesame

Transformation style	Transformation receptor	Vector or gene	Transformation result	References
Agrobacterium-mediated transformation	Seedling stem	A carrot calmodulin gene, *cam-4*	Overexpression of *cam-4* gene enhanced the biosynthetic activity of phenylpropane derivatives	Mitsuma et al. (2004)
	De-embryonated cotyledon	pCAMBIA2301	Regeneration frequency and transformation efficiency were 57.33 and 42.66%, respectively. Several homozygous T_1 plants obtained	Chowdhury et al. (2014)
Agrobacterium tumefaciens-mediated transformation	Cotyledon	LBA4404/pBI121	No transgenic shoots obtained	Taşkin et al. (1999)
Agrobacterium rhizogenes-mediated transformation	Wounded hypocotyl	vector pMOG413	Phytase gene was transferred into the hairy roots	Jin et al. (2005)
	Cotyledon	pCAMBIA2301	Transformation frequency was 1.01%	Yadav et al. (2010)
Floral dipping, pollen infiltration, and suspension drop	Flower	A binary vector containing NPT II gene and EGFP gene	Transformation frequency varied from 3.4 to 50%	Were (2006)
Pollen-tube channel transformation	Jiheizhi No. 1	Insulin gene	Insulin gene was integrated and inherited into T_5 generation	Li et al. (2016)
Biolistic particle gun bombardment	cotyledon	pTABNBAR65 vector contained TA29-Barnase	11 BASTA-resistant plantlets were obtained	Chen et al. (1996)
	5-day-old apical meristem tissues	Bialaphos resistance gene (bar)	T_1 plants obtained	Bhattacharyya et al. (2015)

shoot development stage. The transgenic adventitious shoots were developed based on somatic embryo induction and differentiation techniques. To screen the positive transgenic lines, the optimized medium containing 50 mg·L^{-1} kanamycin and 500 mg·L^{-1} cefotaxime was applied in the study. As a result, the highest plantlet regeneration rate and transformation efficiency in sesame reached 52.00% and 42.66%, respectively. GUS histochemical activity and RT-PCR assay results exhibited the high efficiency of the *Agrobacterium*-mediated transformation technique with low copy of the exogenous gene *npt*II.

To data, there are two successful examples for the application of the *Agrobacterium*-mediated gene transformation in sesame have been reported (Mitsuma et al. 2004; Chowdhury et al. 2017). One is the transformation of a carrot calmodulin gene (*cam-4*) into the wild species *S. schinzianum* ($2n = 64$) through the *A. tumefaciens* infection on the stems of seedlings (Mitsuma et al. 2004). Both southern hybridization and RT-PCR proved the integration and expression of *cam-4* gene in the transformed *Sesamum* plants. Moreover, overexpression of *cam-4* gene enhanced the biosynthetic activity of phenylpropane derivatives in the transformed plants (Mitsuma et al. 2004). The other is the transformation of osmotin-like protein (OLP) gene (*SindOLP*) into the cultivated sesame cv. VR1-1

(Chowdhury et al. 2014, 2017). A total of 13 T0 lines carried single copy of *SindOLP* gene were obtained under the kanamycin screening.

6.4.3 Pollen-Tube Channel Transformation in Sesame

Compared with the above two methods for plant gene transformation, the pollen-tube channel method is relatively simple with low cost, regardless of the tissue culture stage. For sesame, the pollen-tube channel transformation method is rarely used. Zhen et al. (2004) performed the transformation of the exogenous DNA from the wild sesame and sorghum into the cultivated sesame via pollen-tube channel method. To establish a novel plant bioreactor based on sesame oil body expression system, Li et al. (2016) transformed the binary vector PBI121 containing the insulin gene into the sesame cultivar Jiheizhi no. 1 via pollen-tube pathway. The transformed plantlets with kanamycin resistance were testified by PCR and PCR-southern hybridization. The results showed that the insulin gene was integrated into the genome of transformants and passed to the T_5 generation. The findings proved the feasibility of the genetic transformation of the exogenous insulin gene into sesame oil body via the pollen-tube-mediated method. In addition, Were (2006) transformed the green fluorescent protein (*GFP*) reporter gene and neomycin phosphotransferase (*NPT*II) gene into sesame genome via floral dip with *Agrobacterium* suspension before pollination. The highest mean transformation rate

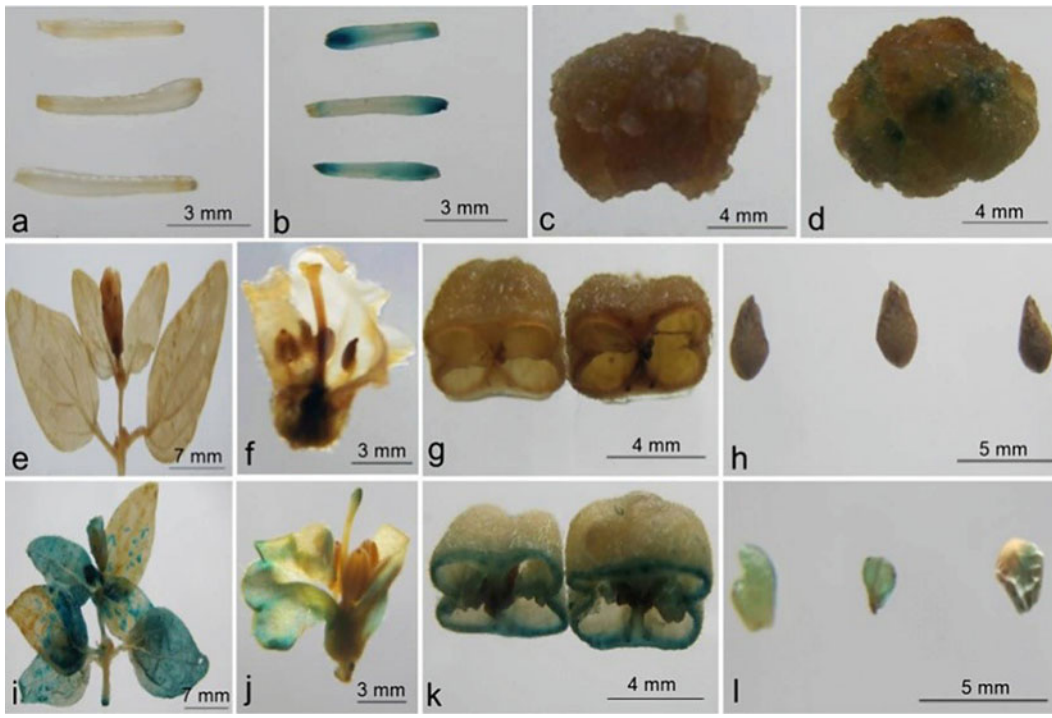


Fig. 6.1 Histochemical assay of positive transformants with transient and stable *GUS* expression. **a** Nontransformed hypocotyl explant (control). **b** Transient *gus* expression in hypocotyl. **c** Nontransformed callus tissue (control). **d** Transient *gus* expression in callus. **e–h** Leaf, flower, capsule, and seed, respectively in the nontransformed plantlet (control). **i–l** Stable *gus* expression of leaf, flower, capsule, and seed, respectively in the positive T0 transformant. (Provided by Ming Ju)

reached 49.8%. The above rare results showed the high potential of uncommon dipping transformation method in sesame.

6.5 T-DNA Insertion Mutant Library in Sesame

In recent few years, in order to improve the genetic modification techniques in sesame, a highly efficient callus induction and differentiation technique and the adventitious shoot induction technique were systematically developed in sesame and several wild species (An et al. 2011; Miao et al. 2012). Subsequently, an efficient *Agrobacterium*-mediated transformation technique via callus induction and differentiation was established in sesame, *S. latifolium*, and *S. radiatum* for the first time (partial results shown in Fig. 6.1) (unpublished data, Haiyang Zhang). The results showed that both callus type and genotype significantly affected the transformation efficiency (unpublished data, Haiyang Zhang).

In the study, the pBI121 vector containing a neomycin phosphotransferase gene (*Npt*-II) and a β-glucuronidase (*Uid* A) gene was transferred into sesame and several wild *Sesamum* species using the transformation system. The embryonic calli (type III, annotated by Miao et al. (2012)) with high quality were induced from the *A. tumefaciens*-infected explants on the selection medium (i.e., basal MS + 0.1 $mg{\cdot}L^{-1}$ NAA + 2 $mg{\cdot}L^{-1}$ 6-BA + 30 $g{\cdot}L^{-1}$ sucrose + 400 $mg{\cdot}L^{-1}$ cefotaxime + 25 $mg{\cdot}L^{-1}$ kanamycin). The histochemical activity of GUS expression proved the positive transformants (Fig. 6.1). As a result, 353 transgenic lines for *Sesamum* were generated. The high transformation efficiency reached to 7.5%. In addition, RT-PCR results indicated that 59.7% of the assayed transgenic lines presented the low copy number of the disarmed T-DNA insertions (≤ 2) (Unpublished data, Haiyang Zhang). The results proved the high efficiency of the *Agrobacterium*-mediated transformation technique in sesame and its key function in genetic transformation and functional genomics research in sesame.

In order to explore the unknown genes related to agronomic traits in sesame, Henan Sesame Research Center, Henan Academy of Agricultural Sciences (HSRC, HAAS), China, recently established a sesame T-DNA insertion mutant library for the first time based on the above techniques (Unpublished data, Haiyang Zhang). Till now, the mutation library consists of 520 mutant lines. PCR assay and phenotypic mutation investigation results showed that the mutation frequency of T-DNA insertion in the library reaches to 2.4%. About 50% of mutated traits were stably inherited into progeny. Meanwhile, the mutated traits were involved in leaf type, stem strength, capsule size, and environmental tolerance and resistance.

Furthermore, to explore the candidate genes affected by T-DNA insertion, the genetic characterization of a subset of 28 positive transgenic T0 lines were analyzed using the new-generation genome re-sequencing technique (Unpublished data, Haiyang Zhang). The results indicated that the average insert number was 1.7. About 54% of the mutated lines presented one inserted site in the whole genome sequences. A total of 21 insertion sites of T-DNA fragment were found located into the 5' or 3' UTR regions of genes, nine in gene exons, seven in gene introns, four in 3' UTR region, and eight in intergenic regions. The high-efficient genetic transformation techniques and the constructed T-DNA insertion library supply the solid basis for the molecular genetics research in sesame.

References

Abdellatef E, Ahmed MMM, Daffalla HM, Khalafalla MM (2010) Enhancement of adventitious shoot regeneration in sesame (*Sesamum indicum* L.) cultivar promo KY using ethylene inhibitors. J Phytol 2:61–67

Ahmed MMM, Abdellatef E, Khalafalla MM (2008) In vitro multiple shoot induction and plant regeneration in elite Sudanese sesame cultivars (*Sesamum indicum* L.). Amer-Euras J Sustain Agri 2:308–314

An JW, Miao HM, Wei SL, Guo WZ, Zhang HY (2011) Study on Induced factors affecting adventitious bud formation from cotyledon in sesame. J Henan Agric Sci 10:44–47

An (2009) adventitious shoot induction and plant regeneration from cotyledon of sesame (*Sesamum indicum* L.). Thesis, Nanjing agriculture university, Nanjing, China

Bangaremma SW, Lokesha R, Gayatree GS, Sharanamma, (2013) Sesame shoot regeneration using different combinations of growth regulators. Mol Plant Breed 4:267–269

Baskaran P, Jayabalan N (2006) In vitro mass propagation and diverse callus orientation on *Sesamum indicum* L.- an important oil plant. J Agri Technol 2:259–269

Bhattacharyya J, Chakraborty A, Mitra J, Chakraborty S, Pradhan S et al (2015) Genetic transformation of cultivated sesame (*Sesamum indicum* L. *cv Rama*) through particle bombardment using 5-day-old apical, meristematic tissues of germinating seedlings. Plant Cell Tiss Org Cult 123:455–466

Breitler JC, Labeyrie A, Meynard D, Legavre T, Guiderdoni E (2002) Efficient microprojectile bombardment-mediated transformation of rice using gene cassettes. Theor Appl Genet 104:709–719

Charaborti P, Ghosh A (2009) Variation in callus induction and root-shoot bud formation depend on seed coat of sesame genotypes. Res J Bot 5:14–19

Chattopadhyaya B, Banerjee J, Basu A, Sen SK, Maiti MK (2010) Shoot induction and regeneration using internodal transverse thin cell layer culture in *Sesamum indicum* L. Plant Biotechnol Rep 4:173–178

Chen ZK, Wang JL, Zhi YB, Yi ML (1994) Study of multiple bud induction and shoot regeneration in sesame. Henan Agric Sci 11:10–12

Chen Z, Zhi Y, Yi M, Wang J, Liang X et al (1996) Transformation of engineered male sterile gene and establishment of transgenic plants in sesame (*Sesamum indicum* L.). Acta Agri Bor-Sin 11(4):33–38

Chowdhury S, Basu A, Kundu S (2014) A new high-frequency *Agrobacterium*-mediated transformation technique for *Sesamum indicum* L. using de-embryonated cotyledon as explant. Protoplasma 251:1–16

Chowdhury S, Basu A, Kundu S (2017) Overexpression of a new osmotin-Like protein gene (SindOLP) confers tolerance against biotic and abiotic stresses in sesame. Front Plant Sci 8:410

Christou P (1995) Strategies for variety-independent genetic transformation of important cereals, legumes and woody species utilizing particle bombardment. Euphytica 85:13–27

Cui YH, Sun Y, Liu WP (2010) Study on in vitro culture of hypocotyl of black sesame. Chin J Guizhou Sci 28:67–72

Gangopadhyay G, Poddar R, Gupta S (1998) Micropropagation of sesame (*Sesamum indicum*) by in vitro multiple shoot production from nodal explants. Phytomorphology 48:83–90

George L, Bapat VA, Rao PS (1987) *In vitro* multiplication of sesame (*Sesamum indicum*) through tissue culture. Ann Bot-London 60:17–21

George L, Bapat VA, Rao PS (1989) Plant regeneration in vitro, in different cultivars of sesame (*Sesamum indicum, L.*). Proc Plant Sci 99:135–137

Honnale HN (2011) Callus induction and organogenesis in *Sesamum indicum* L. CV E8. Curr Trends Biotechnol Pharmacy 6:1330–1332

Jin UH, Chun JA, Han MO, Li JW, Yi YB et al (2005) Sesame hairy root cultures for extra-cellular production of a recombinant fungal phytase. Process Biochem 40:3754–3762

Joshi AB (1961) Sesamum. Indian Central Oilseed Committee Hyderabad, India, p 109

Kwon TH, Abe T, Sasahara T (1993) Efficient callus induction and plant regeneration in *Sesamum* species. Plant Tissue Cult Lett 10:260–266

Li B, Chen R (1990) Tissue culture and plantlet regeneration of sesame. J Shanxi Univ 13:307–310

Li M, Zhang J, Wang Q, Yang YW, Wang ZS (1996) Callus induction and organogenesis of *Sesamum indicum* L. hypocotyl. J Henan Normal Uni 1:61–64

Li X, An SJ, Shao TM, Chai XQ, Liu P et al (2016) Insulin gene transformation into sesame via pollen-tube pathway. Chin J Jiangsu J Agr Sci 32(5):1013–1017

Liu M (1986) Preliminary report on tissue culture experiment of sesame. Chin J Henan Agric Sci 2:15–16

Liu M (1990) Anther culture of sesame (*Sesamum indicum* L.). Henan Sci 2:55–61

Liu SL, Wang YD, Luo S, Zhang J, Han BW (1994) Somatic embryogenesis from hypocotyl culture of sesame and changes of endogenous hormones and soluble proteins during embryogenesis. J Agric Biotechnol 2:44–49

Lokesha R, Shashidhara N, Janagoudar BS (2007) Callus induction and plant regeneration in sesame (*Sesamum indicum* L.) through direct seeding. Plant Cell Biotechnol Mol Biol 8:85–88

Lokesha R, Rahaminsab J, Ranganatha ARG, Dharmataj P (2012) Whole plant regeneration via adventitious shoot formation from de-embryonated cotyledon explants of sesame (*Sesamum indicum* L.). World J Nuclear Sci Technol 2:47–51

Malaghan S, Lokesha R, Savitha R, Ranganatha A (2013) Adventitious shoot regeneration in Sesame (*Sesamum indicum* L.) (Pedaliaceae) via deembryonated cotyledonary explants. Res J Biol 1:31–35

Malaghan S, Lokesha R, Revadi S (2016) Whole plant regeneration through direct organogenesis in sesame (*Sesamum indicum* L.). Green Farming 7:1034–1039

Mary RJ, Jayabalan N (1997) Influence of growth regulators on somatic embryogenesis in sesame. Plant Cell Tiss Org Cult 49:67–70

Miao HM, Ju M, Wei LB, Ma Q, Zhang HY (2012) Establishment of sesame callus induction and shoot regeneration system. Chin Bull Bot 47:162-170

Mitsuma S, Ishigaki E, Sugiyama R, Tetsuya A, Kuoji Y et al (2004) Activation of phenylpropanoid

metabolism in sesame by over-expression of carrot calmodulin gene. Biol Pharm Bull 27:1621–1625

Mukherji S (1947) Relation of total soluble solids in the cell sap of *Sesamum* species to the degree of susceptibility and resistance to *Antigastra* (Lepidoptera–Pyralidæ) attack. Nature 160:95–96

Nimmakayala P, Perumal R, Mulpuri S, Reddy UK (2011) Sesamum. In: Kole C (ed) Wild crop relatives: genomic and breeding resources. Springer, Berlin, Heidelberg, pp 261–273

Pratik P, Shilpa B, Kokiladevi E (2016) Efficacy of plant growth hormones for shoot induction and regeneration in Sesame (*Sesamum indicum L.*). Res J Biotechnol 11 (10):27–30

Qu Z, Wu XY, Xia FJ (1994) Embryo rescue and plant regeneration of inerspecific hybridization between *S. indicum* and a wild specie *S. schizinixianum*. Chin J Oil Crops 1:33–35

Raja A, Jayabalan N (2010) Callus induction and plantlet regeneration from leaf explants of Sesame (*Sesamumindicum L. cv. SVPR - 1*). J Swamy Bot Club 27:93–98

Raja A, Jayabalan N (2011) *In vitro* shoot regeneration and flowering of sesame (*Sesamum indicum* L.) cv. SVPR-1. J Agri Technol 7:1089–1096

Rao KR, Vaidyanath K (1997) Induction of multiple shoots from seedling shoot tips of different varieties of *Sesamum*. Indian J Plant Physiol 2:257–261

Saravanan S, Nadarajan N (2005) Effect of media supplements on *in vitro* response of sesame (*Sesamum indicum* L.) genotypes. Res J Agric Biol Sci 1:98–100

Seo HY, Kim YJ, Park TI, Kim HS, Yun SJ et al (2007) High-frequency plant regeneration via adventitious shoot formation from deembryonated cotyledon explants of *Sesamum indicum* L. Vitro Cell Dev Biol 43:209–214

Shashidhara N, Santosh D, Ravikumar H, Ashoka N, Dhanalakshmi T et al (2011) Exogeneous and endogeneous contaminations in sesame tissue culture—Boon or Bane. Intl J Agric Environ Biotechnol 4:103–106

Shi SW, Cai M (1986) Presentation of plant regeneration of the stem segment culture in wild sesame Congo. Chi J Oil Crops 4:67–68

Shi SW, Cai M (1989) Callus induction and plantlet regeneration from hypocotyls of *Sesamum indicum*. J Plant Physiol 04:40

Shi SW, Cai M (1989) Effect on inducing factors on pollen callus formation of in vitro cultured anthers in sesame (*Sesamum indicum* L.). Chin J Oil Crops 2:45–49

Singh KD, Shagufta K (2011) In vitro regeneration of sesame (*Sesamum indicum* L.) – —an important medicinal oil crop. Crop Res 42(1 to 3):125–130

Singh RP, Singh SP, Pransad BK, Singh BD (2006) Multiple plantlets regeneration in tissue culture of sesame (*Sesamum indicum* L.). Res Crops 7:760–764

Subramanian M (2003) Wide crosses and chromosome behavior in Sesamum. Madras Agric J 90:1–15

Sudhaker D, SreeRangasamy SR, Liang GQ (1990) Sesame seed embryo culture. Chin J Oil Crops 3:102

Taskin KM, Turgut K (1997) In vitro regeneration of Sesame (*Sesamum indicum* L.) Turk J Bot 21:15–18

Taskin KM, Ercan AG, Turgut K (1999) *Agrobacterium tumefaciens*-mediated transformation of sesame (*Sesamum indicum* L.). Turk J Bot 23:291–296

Tiwari S, Kumar S, Gontia I (2011) Mini review: biotechnological approaches for sesame (*Sesamum indicum* L.) and niger (*Guizotia abyssinica* L. f. Cass). Asia Pac J Mol Biol Biotechnol 19:2–9

Wadeyar BS (2013) Sesame shoot regeneration-using different combinations of growth regulators. Mol Plant Breed 4:267–269

Wadeyar BS, Lokesha R (2011) Studies on high frequency shoot regeneration in sesame (*Sesamum indicum* L.). Plant Tiss Cult Biotechnol 21:45–52

Wei SL, Zhang HY, Zheng YZ, Zhang TD, Mei HX et al (2007) Study on callus induction in cotyledon and hypocotyl of sesame. J Henan Agric Sci 2:41–45

Were BA (2006) Genetic improvement of oil quality in sesame (*Sesamum indicum* L.). Thesis, Swedish University of Agricultural Science, Uppsala, Swedish

Were BA, Gudu S, Onkware AO, Carlsson AS, Welander M (2006) *In vitro* regeneration of sesame (*Sesamum indicum* L.) from seedling cotyledon and hypocotyl explants. Plant Cell Tiss Org Cult 85:235–239

Xu ZQ, Jia JF, Hu ZD (1997) Somatic Embryogenesis in *Sesamum indicum* L. cv. Nigrum. J Plant Physiol 150:755–758

Yadav M, Chaudhary D, Sainger M, Jaiwal PK (2010) Agrobacterium tumefaciens-mediated genetic transformation of sesame (*Sesamum indicum* L.). Plant Cell Tiss Organ Cult 103:377–386

Yang M, Liu H, Zhou T, Qu H, Yang Y et al (2017) Production and identification of F1 interspecific hybrid between *Sesamum indicum* and wild relative *S. indicatum*. Sci Agri Sin 50(10):1763–1771

Yi Y, Zhang H, Zuo T, Wang ZZ (1997) Study on the culture of different explants of black seedcoat seame *in vitro*. Acta Agri Bor-Occiden Sin 6(4):26–29

Yifter M, Sbhatu DB, Mekbib F, Abraha E (2013) *In vitro* regeneration of four Ethiopian varieties of sesame (*Sesamum indicum* L.) using anther culture. Asian J Plant Sci 12:214–218

Yu ZG (2005) Study on callus inducement and differentiation of black sesame seed. J Yangtze Univ Agric Sci 2:73–75

Zhang HY, Miao HM, Zhang TD, Wei LB, Li C et al (2013) Biological characters of interspecific hybrid progenies between *Sesamum indicum* L. and wild relatives (*Sesamum schinzianum* Asch, *Sesamum radiatum* Schum & Thonn). Sci Agri Sin 46 (19):3965–3977

Zhang HY, Miao HM, Ju M (2019) Potential for adaptation to climate change through genomic breeding in sesame. In: Kole C (ed) Genomic designing of climate-smart oilseed crops. Springer, Cham, Switzerland, pp 374–376

Zhao RH, Miao HM, Ma Q, Chen CB, Song WQ et al (2018) Karyotype comparison analysis of the wild species *Sesamum alatum* and the cultivated sesame. Acta Scientiarum Naturalium Universitatis Nankaiensis 5(05):27–36

Zhen ZG, Duan Y, Wang XL, Cui XH, Sun MY et al (2004) Breeding of sesame lines with high resistance introduced with foreign DNA by pollen tubes path. Chin J Oil Crop Sci 26:31–34

Zhi YB, Jiang WS, Yi ML, Chen ZK (1998) Study on influence of regeneration frequency of *in vitro* culture of cotyledon of sesame. Chin J Xinyang Agric College 8:12–15

Traditional Breeding in Sesame

7

Haiyang Zhang, D. Ray Langham, Yingzhong Zhao, Ali Khalafalla, and Hongmei Miao

Abstract

S. indicum is the sole cultivated species in the genus *Sesamum*. The relatively low genetic diversity in the cultivated species limits the development of new varieties through breeding. In this chapter, the main breeding methods in sesame including space mutation, heterosis application, hybridization, system breeding, and interspecific hybridization are introduced. Main achievements of the new varieties breeding in China, USA, and African countries are shown in this chapter. The main breeding objectives of the world sesame are discussed accordingly.

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

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

A. Khalafalla
Agriculture Research Corporation, Wad Madani, Sudan

7.1 Introduction

Sesame (*S. indicum*) is the sole cultivated species in the genus *Sesamum*. The narrow genetic base and the relatively low genetic diversity in this species limit the breeding of new elite varieties in sesame. As shown in Chap. 5, mutation techniques and created new lines have been introduced that suggest the possibility of creating abundant breeding materials for sesame genetic improvement. Sesame breeding initiates in the early twentieth century. Pedigree selection and induced mutation are the two most commonly used breeding techniques in sesame (Ashri 2001). Physical or chemical mutagenesis studies started in the 1950s (Shi 1991). In particular, the Plant Breeding and Genetics Section of the Joint FAO/IAEA Division organized the Coordinated Research Project (CRP) in 1993 and encouraged mutation induction and variety improvement in sesame. The CRP project reported 140 useful sesame mutants with new phenotypes related with plant architecture, leaf, flower, capsule, maturation, male sterility, disease resistance, and other agronomic traits developed by the participants by the end of the last century (Ashri 1998, 2001; Van Zanten 2001). Thus, assembling the elite traits such as shatter resistance, determinate, dwarf, and resistance or tolerance to environmental stresses into new sesame varieties became more prevalent and useful. In addition, the traditional breeding techniques including heterosis,

H. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-98098-0_7

space breeding, and pedigree selection and the new varieties developed in sesame are introduced in this chapter.

7.2 Traditional Breeding Techniques in Sesame

7.2.1 Mutation Method and Application in Sesame

Sesame is an ancient oil crop with long cultivation history of more than 5000 years. However, the history of sesame breeding is relatively young. As described in the Chap. 5, physical and chemical mutagenesis has been employed for new germplasm creation in sesame in the past decades. Most of mutant lines have been used for variety breeding and release, such as the Chinese varieties Jiguang 2, Ezhi 2, Jinzhi 3, Zhengzhi 98N09, Ganzhi 9, Yuzhi DS899, and Yuzhi Dw607 (Zhang et al. 2019). To avoid of repetition, we here only introduce the space mutation breeding method and its application in sesame breeding.

Space mutation breeding, also termed space breeding or space technology breeding, is a useful technique for creating new germplasm, new materials, and new varieties in crop breeding. Space mutation was initiated by the first mission of Russia's Sputnik 4 in 1960 carrying the seeds of wheat, pea, maize, and onion. Space mutation breeding is comprised of three steps: seed screening, space mutation, and mutation observation and trait selection. Mutagenized by high-energy ion radiation, microgravity, space magnetic field, ultra vacuum, and other physical stresses in space, the DNAs of seeds or seedlings would be changed and inherited in the progeny (Liu et al. 2004). Space environment could cause up to 2–4% of useful mutations observed in diverse field crops, floricultural plants, and vegetables, compared with their natural variability (1/200,000) and chemical mutagenesis (5/1000) (Pan et al. 2005). In the past five decades, space mutation breeding has been carried out in the USA, the former Soviet Union, Europe, and China. In China, space breeding research started in 1987.

Sesame space breeding was initiated by the 17th retrievable satellite of China in 1996 (Zhang et al. 2003). The seeds of Yuzhi 4 were picked up for 15 d space flight, and the mutated variety Zhongzhi 1 was bred from the progeny of the mutated Yuzhi 4 by the Oil Crops Research Institute of CAAS, China, which should be the first variety bred through space breeding technique. Zhongzhi 1 showed high yield of 1288.8 kg/ha with the increase of 11.82% over the control (Yuzhi 4). The resistance of the variety to charcoal rot disease was also improved at the same time (Zhang et al. 2003). Subsequently, so many sesame varieties such as Zhongzhi 13, Jizhi 1, Jihangzhi 1, Jihangzhi 2, Jihangzhi 3, Jihangzhi 4, and Zhengtaizhi No. 1 were gradually released by Chinese sesame scientists. All the varieties exhibited high seed yield with the increase of 9.55–17.88% over controls, good seed quality, and high resistance to diseases (Zhang 2005; Xu et al. 2010, 2013, 2016; Wei et al. 2016).

Space mutagenesis usually induces morphological variation involved in leaf, plant height, plant characters, flower color, capsule, fertility, pre-flowering period, capsule number per plant, 1000-seed weight, yield, and resistance to environmental stresses in sesame (Lv et al. 2011; Gao et al. 2013). For the mutants, DNA mutagenesis could be identified by amplified fragment length polymorphism (AFLP) markers, as polymorphic DNA bands were diverse and different from the wild types (Sun et al. 2007; Lv et al. 2011; Gao et al. 2013). Considering the space mutation results in the uncertain variation in sesame seeds, a huge mutation bank of sesame should be constructed for new sesame variety breeding and mutagenesis mechanism research of specific traits.

7.2.2 Heterosis Breeding in Sesame

The heterosis phenomenon of sesame was reported by B. Pal for the first time in 1945 (Pal 1945). Sesame has high heterosis potential, the hybrid vigor capacity of the F_1 yield reached up to 252% (Tu 1998; Zheng et al. 2003; Zhang et al. 2005; Sakhiya 2013). In 1993, the first sesame hybrid Yuzhi 9 (ms86-

1 × Danbackggae) was bred and released by the Chinese sesame scientists (Tu et al. 1994). The male sterile line ms86-1 was optimized from an introduced sterile line from the USA by the Chinese sesame scientists and was applied in sesame hybrid breeding. Compared with the control (Yuzhi 4), the yield of the hybrid Yuzhi 9 increased by 29.52% in 1989–1991.

Later, more sterile lines, such as 95 ms^{-5} induced by γ-rays and 0176A and D248A chosen from natural mutants, were detected (Liu et al. 2013; Zhao et al. 2013). The first light insensitive male sterile line *Si*psms304 was created from the sibline of Yuzhi 11 by ethyl methanesulfonate (EMS) mutagenesis in 2009 by the Chinese sesame scientists (Zhang et al. 2016; China patent: ZL201411814406.3). As a result, more hybrids such as Zhengzazhi H03, Zhengzazhi 3, Zhongzhiza 1, Zhongzhiza 2, Wanzazhi 1, Wanzazhi 2, Wanzhi 6, and Wanzhi 11 were gradually bred and released in China. However, most of all the above hybrids are produced based on two-line-producing seed technique. The procedures of the hybrid seed production involve high cost and are time-consuming and thereby limit the application of hybrid seeds in sesame production.

7.2.3 System Selection and Hybridization Breeding in Sesame

In the early of 1950s, sesame scientists began to systematically collect and evaluate the landraces in the world. In China, about 20 landraces, such as Silingcao, Batongbai, Hebei Bawangbian, Baoding Jinkouhuang, Fuyang Silingcao, Wuchang Jiugentou, Xiangyang Laohongzhima, Xiniujiao, Pingyuanbatou, and Daduzhima, were selected and utilized for production in various production regions of China at that time. From 1960 to 1970s, the first-generation Chinese sesame varieties were mainly bred from the elite landrace resources through systematic selection method. As a result, so many important varieties, such as Zhongzhi 1, Xiezhi 1, Xiangzhi 2, Zhuzhi 2, Yuzhi 3, Luozhi 16, Ganzhi 1, Ningzhi 2, and Jinzhi 1 for Hubei, Henan, Hebei, Jiangxi, Jiangsu, and Shanxi provinces, respectively, were released. The varieties played an important role in enhancing the development of sesame breeding in China at early stage.

Hybridization breeding with the pedigree selection is a common and effective method for sesame variety improvement. The desired and wide variation of traits via controlled crosses could be transferred into new varieties. The cross-combinations could be developed between two or more parental lines. In China, the first group of varieties bred through hybridization technique included Zhongzhi 7, Zhuzhi 1, and Jizhi1 that were released in the early 1980s. Later, tens of sesame variety series were bred through hybridization breeding and showed high adaptation for the main production regions in China (such as Henan, Anhui, Hubei, Jiangxi, Hebei, Liaoning, and Shanxi). In the system selection and hybridization breeding, more desirable lines with more elite alleles were rapidly applied for variety improvement. The yield potential and the tolerance to environmental conditions of the representative sesame varieties were improved by the end of twentieth century in China.

7.2.4 Interspecific Hybridization in Sesame

For sesame, interspecific hybridization belongs to resistance breeding. As shown in the Chap. 2, the wild *Sesamum* species are precious germplasm as a source of genes conferring high tolerance and resistance to various pests and diseases and harsh environments. Especially, to our knowledge, there are no varieties or germpalsm accessions immune to the *Fusarium* wilt pathogen *Fusarium oxysporum* and the charcoal rot disease pathogen *Macrophomina phaseolina* except for some wild *Sesamum* species (Zhang et al. 2019). Transferring the donor resistance genes into the cultivated species via interspecific hybridization should be a feasible method to improve sesame varieties (Ashri 2007; Zhang et al. 2013). However, the incompatibility of the different *Sesamum* species under natural condition limits the interspecific hybridization method in sesame breeding (Zhang

et al. 2013; Yang et al. 2017). Thus, before performing the interspecies breeding, the abortion of hybrid embryos and the sterility of interspecies crossing should be overcome. With the aid of immature embryo culture technique, Zhang et al. (2013) obtained the 2430 F_1 interspecific hybrid progeny derived from '*S. schinzianum* × *S. indicum*' and '*S. radiatum* × *S. indicum*' crosses and detected the simple sequence repeat (SSR) marker HS209 to identify the progeny of interspecific hybrids for the first time. The new materials and precise evaluation techniques provide new hopes for interspecific hybridization breeding in sesame, even though no varieties have been bred in sesame using the interspecific hybridization and pedigree selection and proved by molecular markers in the world.

The achievement of the Sesame Genome Project (SGP) gives sesame scientists more chances to apply the elite traits and genetic resources of the wild *Sesamum* species in sesame breeding. Recently, the first interspecific hybrid population for *S. calycinum* ($2n = 32$, var. KEN8) and *S. angustifolium* ($2n = 32$, var. G01) was generated by the Sesame Genome Working Group (SGWG) and applied to construct the first couple of interspecific single nucleotide polymorphism (SNP) genetic maps in sesame (Unpublished data, Haiyang Zhang) (partial results shown in the following chapters). Further molecular mechanism of the elite traits in the wild *Sesamum* species will certainly stimulate the interspecific hybridization and molecular breeding for disease resistance in sesame.

7.3 Sesame Breeding and Representative Sesame Varieties in China

7.3.1 Sesame Breeding History in China

In China, sesame breeding research initiated in the 1950s. Till now at least 181 Chinese sesame varieties including 154 white seeded, 20 black seeded, and 7 yellow or brown seeded varieties have been bred and released for production. Hybridization with pedigree selection is the main classical technique to breed new sesame varieties. From 1950s to 2013, 78 (50.6%) of the 145 varieties were bred by hybridization breeding, followed by system selection breeding (37, 24.0%), mutation induction (11, 7.14%), genetic male sterile (GMS) hybrid (9, 5.8%), and others (20, 20.1%) (Liu et al. 2015). According to the development of breeding techniques, the history of sesame breeding research in China was divided into five stages: (1) 1950–1960s for selection breeding. Dozens of landraces such as Beijing Bawangbian, Shangcai Zihuaye Ersan, Xiangyang Xiniujiao, and Wujinghei were selected from the germplasm library. (2) 1970–1980s for hybridization breeding. Twenty-three varieties such as Zhongzhi 1, Zhongzhi 2, Zhongzhi 3, Zhuzhi 1, Zhengzhi 2, Jizhi 1, and Jizhi 2 are bred. (3) 1990s for mutation, GMS hybrid, and hybridization. Nineteen varieties such as Zhongzhi 5, Zhongzhi 7, Zhongzhi 8, Yuzhi 1, Yuzhi 2, Yuzhi 3, Yuzhi 9, Henan 1, Jizhi 3, and Ningzhi 1 were bred. (4) 2000–2013 for mutation, hybridization, and other breeding techniques. A total of 112 varieties including the Zhengzhi, Ganzhi, Zhongzhi, Ezhi, Wanzhi, Jinzhi, Liaozhi, and Jizhi series were bred and released. (5) 2014 to present for mutation, hybridization, and molecular breeding techniques. Besides traditional variety breeding, some new varieties with new elite traits, such as Yuzhi DS899 with determinacy, Yuzhi Dw607 with short internode length and plant height, and their derived varieties, are bred and released. At present, the new lines with shattering resistance, indehiscence, and high oleic acid content have been created and applied for variety improvement in China (Data not shown). Thus, we believe that the new progress in sesame breeding techniques and variety improvement would stimulate the sesame industry in near future.

7.3.2 Parental Lines in Sesame Breeding in China

In China, at least 8115 sesame germplasm accessions are preserved to date, which cover the

46 countries in the world. More than hundreds of elite germplasm accessions have been utilized in breeding (http://www.sesame-bioinfo.org/phenotype/index.html) (Yang et al. 2018). Pedigree analysis revealed that Yuzhi 4, Ezhi 1 (old), and Yiyangbai are the most important breeding materials as parental lines for sesame breeding in China (Yang et al. 2018). At least 21 varieties are bred using Yuzhi 4 as the parental line. To breed an elite variety, more outstanding parental lines with different blood relationship and geographical origins are required. At present, in order to enhance the selection efficiency, most sesame varieties in China are derived from the 45 Chinese original local resources and nine foreign resources using hybridization. The high-efficient application of the parental lines could rapidly realize the objectives of sesame breeders. However, we should know that the loss of amount of alleles in landraces and narrower genetic basis become more and more serious. The challenges of creation of new materials and highly efficient breeding techniques for sesame breeding should be faced and solved in the future, in order to change the slow breeding status.

7.3.3 Representative Sesame Varieties in China

Seed yield and seed quality are the two key traits in sesame breeding. From 2011 to 2015, 39 sesame varieties were bred and released in China (Table 7.1). As shown, the yield level of the new varieties increased to 120.6 kg per mu, equal to 1809 kg per ha (in Yuzhi Dw607). The highest oil and protein content of varieties reached to 60.52% (in Zhongzhi 12) and 26.75% (in Yuzhi Ds899), respectively. In sesame new variety examination in China, the standard of good sesame variety comprise of the six items, i.e., (1) seed type and color: oval and pure white or black seed; (2) ratio of seed coat weight: $\leq 10\%$; (3) 1000-seed weight (TSW): >2.5 g; (4) seed oil and protein content: Oil content $\geq$ 54% or protein content $\geq 20\%$; (5) total sugar content of seed: $\geq 10\%$; (6) total lignan content of seeds: 0.5%. As to the varieties with high oil content (>58%), high protein content (>24%), large size (TSW > 4.0 g), or tiny size (TSW < 2.0 g) are regarded as high-quality variety.

Yuzhi 11

Yuzhi 11 (Fig. 7.1) is a famous and popular sesame variety in China, which was bred using mutation induction method by the Henan Sesame Research Center, HAAS in 2003, which has been applied for genome sequencing in the SGP. Yuzhi 11 has high and stable yield with the record of 3,204 kg per ha in summer sowing in Henan in 2004. The seeds contain 56.66% fatty acids and 19.74% protein.

Ganzhi No. 9

Ganzhi No. 9 (Fig. 7.2) is the first Chinese black sesame variety bred through radiation mutation in China. The variety has high yield and commodity traits with high resistance to *Macrophomina phaseolina* and high tolerance to lodging and drought.

Yuzhi DS899

Yuzhi DS899 is the first determinate variety for production application in the world (Fig. 7.3). The variety was bred and released in 2015. Yuzhi DS899 was mutated from the sibline of Yuzhi 11 using EMS mutagenesis. The variety has high yield potential of 130 kg per mu with high density of 15,000–18,000 plantlets per mu. The oil content and protein content of seeds is 52.21 and 26.05%, respectively. The thousand seed weight (TSW) is 5.1 g. Thus the variety is regarded as high-quality sesame variety and suitable for direct seed processing.

Yuzhi Dw607

Yuzhi Dw607 is a dwarf variety with short internode length. The variety was bred and released in 2015 (Fig. 7.4). Yuzhi Dw607 was mutated from the sibline of Yuzhi 11 using EMS mutagenesis. The internode length is 3.8–4.6 cm. The plant height ranges from 1.1 to 1.6 m. The variety is suitable for high density cultivation and mechanized harvest with abundant fertilizers and water. For the low plant height and strong stem, the resistance to lodging of the variety is high.

Table 7.1 New sesame varieties released by Chinese sesame scientists from 2011 to 2015

Year	Breeding organization	Variety name	Average yield (kg/mu)	Increase rate (%)	Oil content (%)	Protein content (%)
2011	Oil crops research institute, CAAS	Zhongzhi 18	59.1	13.22	56.82	19.89
	Oil crops research institute, CAAS	Zhongzhi 19	100.37	12.64	**60.52**	–
	Oil crops research institute, CAAS	Zhongxiangzhi 1	84.63	31.23	59.17	–
	ZAAS, Henan	Zhuzhi 19	84.15	7.57	54.71	21.96
	HAAS, Hebei	Jiheizhi 2	72.25	23.60	52.12	24.15
	AAAS, Anhui	Wanzhi 5	92.2	8.14	56.38	–
	SAAS, Shanxi	Fenzhi 7	65.51	11.35	53.80	21.99
2012	HSRC, HAAS, Henan	Zhengzhi 15	79.24	14.55	58.66	16.70
	HSRC, HAAS, Henan	Zhengzhi 16	90.02	6.31	59.04	16.80
	HSRC, HAAS, Henan	Zhengzhi 17	94.04	11.03	58.34	17.96
	Oil crops research institute, CAAS	Zhongzhi 22	71.77	19.0	58.67	18.23
	Oil crops research institute, CAAS	Zhongzhi 23	77.02	5.58	55.61	20.90
	Oil crops research institute, CAAS	Zhongzhi 21	79.48	14.16	56.95	18.91
	Oil crops research institute, CAAS	Yanzhuang 1	90.69	7.09	59.27	18.70
	AAAS, Anhui	Wanzhi 6	91.69	8.27	57.67	–
	SAAS, Shanxi	Jinzhi 7	76.90	9.42	55.40	22.52
	LAAS, Henan	Luozhi 21	78.94	5.02	55.69	21.25
	XAAS, Hubei	Ezhi 7	80.19	15.18	56.71	18.85
2013	HSRC, HAAS, Henan	Zhengzhi 18	88.0	11.60	56.71	–
	Oil crops research institute, CAAS	Zhongzhi 24	88.7	12.24	52.70	16.66
	Oil crops research institute, CAAS	Yanzhuang 2	97.7	15.35	–	–
	AAAS, Anhui	Wanzhi 7	89.5	5.51	58.32	–
2014	HSRC, HAAS, Henan	Zhengzhi 19	108.2	10.21	55.0	21.0
	HSRC, HAAS, Henan	Zhengtaizhi 1	115.5	17.69	57.49	17.49
	JAAS, Jiangxi	Ganzhi 11	78.74	10.13	51.28	**25.53**
	JAAS, Jiangxi	Ganzhi 12	94.43	7.27	57.06	20.57
	JAAS, Jiangxi	Ganzhi 13	95.81	8.84	57.48	21.82
	Oil crops research institute, CAAS	Zhongzhi 29	105.7	7.82	–	–
	HAAS, Hebei	Jihangzhi 3	78.07	5.57	53.45	22.9
	ZAAS, Henan	Zhuzhi 21	87.30	11.80	56.0	20.3
	XAAS, Hubei	Ezhi 8	81.36	7.17	56.53	19.76

(continued)

Table 7.1 (continued)

Year	Breeding organization	Variety name	Average yield (kg/mu)	Increase rate (%)	Oil content (%)	Protein content (%)
2015	HSRC, HAAS, Henan	Yuzhi DS899	117.7	9.62	**51.43**	**26.75**
	HSRC, HAAS, Henan	Yuzhi 21	112.6	14.9	58.33	17.92
	HSRC, HAAS, Henan	Yuzhi Dw607	120.6	12.3	52.62	22.28
	NAAS, Henan	Wanzhi 16	110.1	12.3	56.94	19.21
	ZAAS, Henan	Zhuzhi 22	106.8	8.9	52.74	19.04
	NAAS, Henan	Wanzhi 16	110.1	12.3	56.94	19.21
	ZAAS, Henan	Zhuzhi 22	106.8%	8.9	52.74	19.04
	AAAS, Anhui	Wanzhi 12	–	–	–	–

'–' indicates not known. (Provided by Haiyang Zhang)

Fig. 7.1 Plant and seed image of sesame variety Yuzhi 11. Left is the plant image in field. Right is the image of seeds. Photographs provided by Haiyang Zhang

Fig. 7.2 Images of Ganzhi No. 9. Photographs provided by Meiwang Le

Fig. 7.3 Plant images of Yuzhi DS899. Photographs provided by Haiyang Zhang

Yuzhi Dw607 variety has high yield potential of 160 kg per mu with high density of 14,000–15,000 plantlets per mu. The oil content and protein content of seeds is 52.62 and 22.28%, respectively. In 2017, Yuzhi Dw607 was planted in Xinjaing Experimental Station and created the highest yield record of 233 kg per mu (equal to 3495 kg per ha) in the world.

In recent years, under the support of the Chinese Agriculture Research System, sesame scientists in Henan Sesame Research Center cooperated with the Agriculture Research Corporation in Sudan and performed the sesame breeding experiments in Sudan. At present, three lines (CSS1, CSS2, and CSS3) have been bred with the Chinese and African varieties as parental lines using hybridization breeding technique (Fig. 7.5). In 2018, the highest yield level reaches 103.1 kg per mu with the increase of 108.28%, compared to the control. The selection of the new varieties will improve the sesame industry in Africa (Unpublished data, Haiyang Zhang).

7.4 Sesame Breeding and Representative Sesame Varieties in Africa

7.4.1 Sesame Breeding Development in Africa

Daniel (2017) stated that many constraints such as indeterminate flowering nature, shattering of capsules at maturity, insect, pests and diseases, heat, and drought are the major factors hindering the considerable yield improvement in Ethiopia. The improvement of sesame has lagged behind other crops due to a lack of research, shortage of trained personnel, limited financial support, and limited international cooperation (Daniel 2017). In Sudan, the main problems limiting production and expansion of sesame are the low yield potential of existing varieties, rain fall amount and distribution, limited use of the deficient certified seeds, inadequate weed control, harvest problems mainly related to dehiscence, labor

Fig. 7.4 Image of Yuzhi Dw607. Left is the plant image. Right is the comparison of internode length of Yuzhi Dw607 (right) with the control (left). Photographs provided by Haiyang Zhang

Fig. 7.5 Three new varieties (lines) bred by Chinese sesame scientists in Sudan. The left, middle, and the right image indicate CSS1, CSS2, and CSS3, respectively. Photographs provided by Haiyang Zhang

shortage and too expensive harvest stage, and prevalence of many insect pests and diseases. Lacking of non-shattering varieties and difficulties in mechanical harvesting are of the main concern, whereas the harvesting cost of sesame is more than 70% of the total cost. Khidir (1969) stated that in Sudan as well as in many other countries, the main problem facing sesame production is the shattering of seeds at maturity, thus leading to a great loss in yield.

The main objectives of sesame breeding are high seed yield, indehiscent capsule, high oil percent, and high tolerance to pest and diseases. Sesame improvement resulted in rapid replacement of the old landraces which is characterized by as excellent source of genes for adaptability, and resistance to biotic and a biotic stresses by new ones of narrow genetics background. Due to proper management of the sesame genetic resources should be achieved, this management includes collection of sesame germplasm, conservation, evaluation, and characterization. In Africa, the main breeding methods are utilized to identify and select the sesame varieties possessing desirable traits. Breeding methods used for sesame improvement vary from selection to hybridization, mutation, and molecular breeding. Most of the sesame varieties released in African countries were developed by selection procedure or through hybridization. These methods of breeding are technically easy, need no sophisticated tools, and are suitable for improvement of many traits with unidentified genes at one time.

7.4.2 Breeding for Indehiscent Sesame Varieties in Sudan

In Sudan, the ultimate objective of sesame breeding program, since its inception in the early 1950s, has been the development of high yielding non-shattering varieties for mechanized crop production. In 1951, a segregating population from a cross with an indehiscent mutant was imported from the USA and initial selections were made at the central rain-land research station at Tozi till 1954. The selected lines like the original materials were proved to be poor yielders (Walton 1959). In 1957, crosses were made to transfer the indehiscent trait to the local types. However, the lines resulting from this program were also poor yielders. At the end of 1960s, a multiple crosses program involving several indehiscent lines and a large number of local types started, but no success was achieved (Khidir 1998). It is noteworthy that in many parts of the world breeding for indehiscence proved to be a difficult undertaking. Thus, the main and immediate objective of the sesame breeding program should be to develop the high yielding

but indehiscent varieties. The shattering problem could be partially solved by breeding for uniform maturity (Khidir 1969).

7.4.3 Local Varieties and New Varieties in Sudan

The local sesame varieties in Sudan are numerous with many local names, such as Housh, Abdelrazig, Babinga, Bilia, Abusofa, Heraher, Gerabin light, Gerabin late, Abuzabad, Amortaba, Baladi, and Mafaza light. These varieties are characterized by great variability in the vegetative and reproductive characteristics, as well as the seed color. The seed color varies from white to black with all shades of color in between. The white seeded varieties are preferable for exporting, while the colored for oil extraction.

Landraces are still widely used in Sudan, particularly in Western Sudan. In the 1950s and the early 1960s, a large number of landraces were collected. By selection within these landraces and purification of local varieties according to the seed color, some local varieties, such as Ziraa-1, Ziraa-3, Ziraa-6 and Ziraa-7 in white seeded were released. The varieties were available to the commercial and semi commercial farmers. In 1960s, with the objectives of producing non-shattering high yielding varieties for mechanized production, a population segregating for the indehiscent gene (a recessive mutant discovered in Venezuela in 1943) was imported from the USA. The selected lines from this population and the lines produced by transferring the gene to the local types were proved to be poor yielders.

In 1975, a joint breeding program between Agricultural Research Corporation (ARC) and University of California, Riverside, started with financial support from UNDP. This led to the release of the variety Kenana-1. In mid-1980s, Kenana-1, Ziraa-9, and hurria-49 were released (Osman 1985). Kenana-1 is characterized by high yield productivity, large seed size, and white seeds that contain 48% oil and 24% protein. The variety is resistant to the bacterial blight but attractive to the web worm. The capsule has eight locules. Ziraa-9 is characterized by small white seed and uniform maturity (Osman 1985).

In 1990s, Kenana-2, Kenana-3, and Kenana-4 were released. Kenana-2 is characterized by big seed size with white color. It is also of early maturity and drought tolerant. Promo is a variety selected from introduced materials of temperate origin (Greece), characterized by high branching, medium duration, even maturity and delayed shattering (Ahmed 2008). In 2003, Um Shagara and Gedarif-1 were released. Gedarif-1 is a variety selected from segregating materials of crosses between temperate and tropical cultivars. It is characterized with non-branching habit, medium–late duration to flowering and good vigorous habit of growth. Um Shagra is selected from segregated materials of crosses between introduced and local cultivars. The variety is characterized by high branching, white seed, and even maturity (Ahmed et al. 2003).

In 2012, a new sesame variety named as Elgizouli was released. The variety is selected from segregated materials of crosses between introduced and local cultivars. The original seed stock of the introduced parent was from International Development Research Centre (IDRC) project. It is characterized by high branching, long capsule, white seed, and even maturity (Khalafalla and Ahmed 2012).

7.4.4 Sesame Breeding and New Varieties in Ethiopia

Sesame breeding in Ethiopia started in late 1960s with 72 introduced and landrace accessions, focusing on introduction (Tadele 2005). Ethiopian sesame breeders have come to rely on breeding methods involving the use of collection, selection (mass and pure line), hybridization (followed by pedigree and bulk method of breeding), and recently induced mutation (Daniel 2017). Ethiopian Institute of Agricultural Research (EIAR), The Institute of Agricultural Research (IAR), and IDRC initiated the high land and low land oilcrop projects in 1981 and 1982, respectively. The low land project covered sesame, ground nut, caster, and safflower. The

project continued with its second phase from 1988 and greatly accelerated the research work on sesame. During this time, both exotic and local sesame lines were evaluated for their disease resistance, yield, and yield attributing characters at various ecological zones of Ethiopia. Out of sesame lines assembled and evaluated at multilocations, 10 lines (exotic and some landraces), have been proved promising, and four varieties (i.e., T 85, Kelafo 74, S, and E) were released during the first phase of the project. The cultivar T 85, which is also called Hirhir as local name and Humera type as market name, is appreciated worldwide for its aroma and sweet taste. The variety has good uniform large white seed and often used as standard in the world market. However, it is susceptible to bacterial blight caused by *Pseudomonas* or *Xanthomonas*. Hence, other three varieties S, E (introductions from Uganda) and Abasina (selected line from landraces) were released which meant to mitigate the limitation in T 85, as the varieties moderately tolerate the disease (Tadele 2005).

Daniel (2017) reported that T 85 and Kelafo were released on 1976 and the varieties E and S were released on 1978. Six sesame varieties (i.e., Adi, Mehedo 80, Abasina, Argene, Serkamo, and Tate) were evaluated further and released in the second phase of the project. All of these ten sesame varieties are selected lines from the landraces (Thomas Development Associates 1992). The varieties Mehado8 released on 1989 and Abasina released on 1990. Argene, Adi, and Serkamo were released on 1993, and Tate was released on 2000 (Daniel 2017).

The productivity of sesame is induced by improved varieties or improved agronomy practices and crop protection. The potential yields are probably as high as 2000 kg/ha which is much higher than the national average 600 kg/ha (Wijnands et al. 2009). The data indicate that the yield potential of sesame is much higher than the actual yield, as still much damage occurs by pests and diseases, insufficient weed control, high level of mono cropping, inefficient harvesting for shatter loss, and unrealized genetic potential. The main objective of sesame research in National Agricultural Research System of Ethiopia is to develop new cultivars concentrating on diseases resistance, white seeded, and high yielding characters with increased market values. The research in the country is geared toward evaluation of new collected accessions and exotic materials to develop a sesame variety with desirable characteristics with a combination of traits of interest including considerable oil content and other seed quality traits important for the export market (Tadele 2005).

7.4.5 Sesame Breeding and New Varieties in Tanzania

A large number of Tanzanian farmers are still subsistence producers using the diverse crop cultivars originated from their own farming environments. The number of traditional cultivars selected to adapt for the different cultural values and environmental conditions is also very large. Many traditional cultivars are given names which emphasize the qualities of the cultivars, thus resulting in easy sharing of information and seed of such cultivars. As a result, traditional cultivars are easily accepted and spread within ethnic ground.

The national varieties listed since 1950s up to 2008 in Tanzania include Naliendele92 released in 1992, which is characterized by basal branching and heavy leaf shedding at maturity but susceptible to the bacterial blight. The variety Zawadi94, released in 1994, is characterized by profuse basal branching. The capsules turned from green to purple at maturity. The seed color is dull white. The variety tolerates to leaf spot and stem rot. Ziada94 was released in 1994, characterized by late maturity (120–140 d). The plant color is light green. The variety has profuse branching, the seed color white brown but change to creamy white, if harvested in prolonged rains. The variety Lindi 02 was released in 2006, characterized by basal branching, long capsule, 2–3 capsules within the same plant with very white seeds, and tolerant to the leaf spot and stem rot. The variety has high oil content (55–61%).

7.5 Sesame Breeding in the USA and Ideal Sesame

Since 1950s, main objectives of sesame breeding in the world involves adaptation to mechanization with determinate, dwarf, not-dehiscent or indehiscent trait. With the aid of abundant sesame germplasm materials, D. G. Langham and D. R. Langham performed experiments to breed new sesame varieties with high adaptation for mechanized harvest for more than 70 years in the USA. In 1982, Sesaco Company released varieties that could be swathed into a windrow. In 1988, Sesaco released the first sesame varieties that could be harvested directly in the field after the sesame was dry enough for a 6% seed moisture. In 1997, a non-dehiscent (ND) sesame variety was released in the USA with improved shattering resistance. In 2008, the first improved non-dehiscent (IND) sesame variety (US Patent no. 8,080,707) was released. After being dry for one month, 85% of the capsules in IND plants still have seed to the top (Langham 2008). The development of the first indehiscent sesame variety realizes the whole mechanization production in the USA, even though the seed quality should be improved further. 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). The history of breeding of indehiscent varieties in the USA shows that more non-dehiscent sesame varieties could be developed and spread to other areas of the world. Consequently, full mechanization with non-dehiscent sesame varieties would benefit the world sesame industry in the future.

As breeders, we all must admit that each of us has a set of prejudices as to what sesame should be liked. Just on the basis of three basic traits of branching, capsules per leaf axil, and days to maturity, it is possible to categorize the origin of new germplasm materials: (1) single stem, triple capsules, medium maturity from China; (2) few branches, triple capsules, early maturity from the Republic of Korea; (3) few branches, single capsules, medium-to-early maturity from Turkey and the Middle East; and (4) many branches, single capsule, medium-to-late maturity from America, Africa, and Indian subcontinent. A few other traits further define the latter phenotype: (1) long-day flowering: North America; (2) short-day flowering: the rest of America and Africa; (3) smaller leaves: Indian subcontinent; (4) red seed: South India. Many countries have collected extensive germplasm and then developed a core collection. There must be a concerted effort to not concentrate on just the core collections. There are many hidden genes in the germplasm materials that did not qualify as core. There is a tremendous amount of variation in morphological traits and many more to be discovered with microscopes, chemical analysis, and genomics.

As to the ideal sesame variety, the following are different concepts of ideal sesame from different sesame scientists. In Japan, T. Kobayashi's ideal sesame (Kobayashi 1986) can be described by: non-branching, short internodes, early maturing, low plant height, triple capsule, bicarpellate, opposite leaf phyllotaxy, and lodging resistance.

In Republic of Korea, C. W. Kang's ideal sesame (personal communication, 1998): shatter resistance; white or black seed (brown is good but there is no market); few branches, branches, or uniculm; triple capsule or one capsule; semi-dwarf; dense capsule set; super yield—F_1 hybrid using GCMS; high maturity to the top; clean (clear) color; resistance to Phytophthora blight, *Fusarium* wilt and *Corynespora* blight; low-temperature germinability; high harvest index—low capsule set; lodging resistant; determinate with numerous capsule nodes; qualitative improvement over current varieties: oleic or linoleic fatty acid, oil, sesamin, and sesamolin; taste, protein, essential amino acids.

In Thailand, W. Wongyai's ideal sesame (personal communication, 2000): white seed (major) and black seed, shaker shatter resistance for manual and machine threshing; high seed weight per capsule; 100-seed weight > 3 g; delayed shattering 10–20 days from maturity; early maturity 80–90 days to maturity for early rainy season; medium maturity 95–100 days for

rainy season; suitable seed yield for early maturity; resistance to bacterial leaf spot (*Pseudomonas* and *Xanthomonas*), whitefly, and aphid; having capsule at the tip of stem (may give uniform ripening and good seed fill); short internode, especially look for short internode in the lower part; small leaf size and erect angle and leaf fallen before harvesting; bicarpellate capsule with opposite position and alternate arrangement; single capsule or three capsules which have a uniform capsule size; plant height 120–150 cm; single stem or branching with 2–4 branches but prefer two branches; branches should occur at the same node and parallel to stem; erect stem; lodging resistance; no vivipary; high lignans antioxidant; no or small nectaries (Thai lines have some of the smallest nectaries); non-photosensitive; no longer capsule length; look for later flowering approximately 40 days; shorter ripening period.

In Australia, M. R. Bennett's ideal sesame (personal communication, 2016): some of the characters were mandatory: seed color white to cream; seed should not be bitter; tolerance to diseases, insects, and lodging; minimal shattering; >0.38 g/100 seeds with a minimum of 0.32 g; >44.5% harvest index; 56.1–27.0% oil content, with higher content rated a 4; 40–50 cm height of first capsule with lower values for both higher and lower heights; 2.26–3.00 branches for early and medium maturity and 0.75–1.00 branches for late maturity and anything over 3.0 branches; 120–130 cm plant height with lower values lower and higher; the ideal was 120–130 cm plant height, 40–50 cm first capsule height, 82 cm capsule zone length, with 28 'whorls' (node pairs) of triple capsules based on 20 plants/m^2; and >90 capsules per plant. The following traits were desirable but were not scored: capsule color change at maturity; short time between the end of flowering and maturity; lowest branch should be basal; seedling vigor; dormancy to control vivipary; and capsule length should be longer than internode length. He took data on several traits, but they were not scored: number of nodes to first capsule; days to 50% flowering, days to 90% flower termination, days to 90% mature; capsule length; capsule width; hairiness on the leaf, stem, and capsule; seed weight per plant; and biomass per plant.

In Turkey, H. Baydar's ideal sesame: branches; single capsule; bicarpellate capsule (Baydar 2005).

The above concepts and ideas for the ideal sesame indicate that there is no ideal sesame that will be suitable for the whole world**.** The ideal sesame for Chinese manual farmers is not the same as for USA mechanized farmers. Even within China, the ideal sesame for the north with a shorter growing period will be different than for the south with a longer growing period. Therefore, we breeders should apply the high-efficient breeding techniques involved in the traditional breeding, molecular marker assisted selection, and molecular design breeding techniques to realize the individual ideal breeding objectives in the future.

References

Ahmed ME (2008) Evaluation of new sesame (*Sesamum indicum* L.) genotypes for yield, yield components and stability. Univ Khartoum J Agric Sci 16:380–394

Ahmed ME, MB Taha, Khalafalla AA (2003) Assessment of new promising sesame cultivars for the central rain lands and irrigated systems. Variety Release Committee of Sudan.

Ashri A (1998) Sesame breeding. In: Janick J (ed) Plant breeding reviews, vol 16. Wiley & Sons, Israel, pp 179–228

Ashri A (2001) Induced mutations in sesame breeding. No. IAEA-TECDOC—1195

Ashri A (2007) Sesame (*Sesamum indicum* L.). In: Singh RJ (ed) Genetic resources chromosome engineering, and crop improvement, oilseed crops. CRC Press, Taylor and Francis Group, Boca Raton, US, pp 231–289

Baydar H (2005) Breeding for the improvement of the ideal plant type of sesame. Plant Breed 124(3):263–267

Daniel EG (2017) Sesame (*Sesamum indicum* L.) breeding in Ethiopia. Int J Novel Rese Life S 4(1):1–11

Gao T, Wei S, Li C, Zhang H, Li F et al (2013) Effects of space-induced mutation and analysis of polymorphsim by AFLP molecular marker in sesame. Acta Agri Bor-Sin 1:227–233

Khalafalla AA, Ahmed ME (2012) Proposal for the release of New Elite Sesame Genotypes for Rainfed. A Paper Submitted to the National Variety Release Committee, May 2012, Khartoum, Sudan

Khidir MO (1969) Indehiscent sesame (*Sesamum orientale* L.). Sudan Agric J 4:30–36

Khidir MO (1998) Oilseed crops in the Sudan. (in Arabic), Khartoum University Press, Sudan

Kobayashi T (1986) Early maturing, short internode varieties of sesame. In: Martínez JF (ed) Sesame and safflower newsletter 2. The Center of Agrarian Research and Development, Cordoba, Spain, pp 33–35

Langham DR (2008) Growth and development of sesame. American Sesame Grower Association, San Antonio, TX, p 44

Liu L, Wang J, Zhao L, Yang J, Guo H et al (2004) Research progress in mutation effects of aerospace on crop and ground simulation on aerospace environment factors. Acta Agri Nucleat Sin 18(4):247–251

Liu H, Yang M, Wu K, Zhou X, Zhao Y (2013) Development, inheritance and breeding potential of a recessive genic male sterile line D248A in Sesame (*Sesamum indicum* L.) SpringerPlus 2(1):268

Liu Y, Mei H, Wu K, Zheng Y, Zhang H (2015) Effects of steeping temperature and chemical mutagen on seeds germination of sesame. J Henan Agric Sci 41(11):39–44

Lv H, Zhang Y, Wang L, Wei W, Li D et al (2011) Sesame mutant induced by space flight treatment and analysis of polymorphsim by AFLP molecular marker. J Nuclear Agric Sci 25(5):0851–0856

Osman HE (1985) Sesame growing in Sudan. In: Ashri (ed) Sesame and safflower status and potentials. FAO Plant Prod Prot Paper No 66, Italy, pp 48–51

Pal BP (1945) Studies in hybrid vigour. Indian J Genet 5:106–121

Pan G, Yin X, Yang Q, Zhang Y (2005) Research progress on space-flight breeding of crops. Southwest Hort 33(4):34–36

Sakhiya SV (2013) Diallel analysis in sesame (*Sesamum indicum* L.) Thesis. Junagadh Agricultural University, Junagadh, India

Shi, (1991) Application of mutagenesis technique in the improvement of sesame varieties. Chin Oil Crop 2:93–96

Sun J, Tu Y, Zhang X (2007) DNA fingerprint analysis in space-induced sesame mutant lines. Sci Agri Sin 40 (12):2696–2701

Tadele A (2005) Sesame (*Sesamum indicum* L.) research in Ethiopia: a review of past work and potential and future prospects. In: Martinez JF (ed), Sesame and safflower newsletter. FAO, Cordoba, Spain

Tu L (1998) Research overview on heterosis utilization in sesame. J Henan AgricSci 2:11–13

Tu L, Wang W, Liang X, Zheng Y, Liu J et al. (1994) Breeding and utilization of sesame hybrid Yuzhi 9. J Henan Agric Sci(5):8–10

Van Zanten L (2001) Sesame improvement by induced mutations: Results of the co-ordinated research project and recommendation for future studies. In: Van Zanten L (ed) Sesame improvement by induced mutations 1, Proc Final FAO/IAEA Co-ord Res Mtng IAEA, Vienna, Austria, pp 1–12

Walton PD (1959) Plant breeding in central rainlands of the Sudan. Empire J Exp Agri 27:79–85

Wei S, Gao T, Zhang H, Li F, Li C et al (2016) Breeding of new high-yield superior sesame variety Zhengtaizhi No.1. J Henan Agric Sci 45(12):49–51

Wijnands JHM, Biersteker J, Van Loo EN (2009) Oilseeds business opportunities in Ethiopia. Public private partnership in oil seed, Report, Ministry of Agriculture, Nature and Food Quality, Government of the Netherlands

Xu G, He J, li Y, Cheng Z, Chen S (2010) Breeding of a new sesame variety Jihangzhi 1 and its cultivation technique. J Anhui Agric Sci 29:16180–16181+16187

Xu G, Wang S, Zhu D, Guan Z (2013) Analysis on the high yield, stable yield and adaptability of new sesame variety Jihangzhi No. 2. J Hebei Agric Sci 4:59–61

Xu G, Xu J, Guang Z, Lin G, Guo Y (2016) Breeding and adaptation analysis of new sesame variety Jihangzhi No. 4 with high yield and disease resistance. J Hebei Agric Sci 6:77–80

Yang M, Liu H, Zhou T, Qu H, Yang Y et al. (2017) Production and identification of F1 interspecific hybrid between *Sesamum indicum* and wild relative *S. indicatum*. Sci Agri Sin 50(10):1763–1771

Yang W, Gao Y, Wei X, Wang L, Li D et al (2018) Design and establishment of information database of sesame (*Sesamum indicum* L.) germplasm resources. Chin J Oil Crop Sci 1:57–63

Zhang X (2005) Introduction of Zhongzhi 13. Bull Agric Sci Technol 6:50

Zhang X, Li P, Cheng Y, Feng Y, Fu L (2003) Breeding of sesame variety Hangzhi No. 1 and production techniques thereof. Chin J Oil Crop Sci 25(3):34–37

Zhang TD, Zhang HY, Zheng YZ, Wei SL, Mei HX et al (2005) A review of heterosis and seed production techniques in sesame. Crops 5:64–67

Zhang HY, Miao HM, Zhang TD, Wei LB, Li C et al. (2013) Biological characters of interspecific hybrid progenies between *Sesamum indicum* L. and wild relatives (*Sesamum schinzianum* Asch, *Sesamum radiatum* Schum & Thonn). Sci Agri Sin 46 (19):3965–3977

Zhang H, Miao H, Wei L, Zhang Z, Li C et al. (2016) A sesame light-insensitive male sterile line and the hybrid seed producing method. China patent, ZL. 201411814406.3. Authorization date: 2016.03.08.

Zhang HY, Miao HM, Ju M (2019) Potential for adaptation to climate change through genomic breeding in sesame. In: Kole C (ed) Genomic designing of climate-smart oilseed crops. Springer, Cham, Switzerland, pp 374–376

Zhao Y, Yang M, Wu K, Liu H, Wu J et al. (2013) Characterization and genetic mapping of a novel recessive genic male sterile gene in sesame (*Sesamum indicum* L.). Mol Breed 32(4):901–908

Zheng Y, Zhang H, Mei H, Zhang T, Wei S et al (2003) Hybrid sesame research advances in China. J Henan Agric Sci 11:17–19

Molecular Mapping and Breeding in Sesame

8

Haiyang Zhang, Yingzhong Zhao, Libin Wei, Ting Zhou, Yinghui Duan, and Hongmei Miao

Abstract

Next-generation sequencing (NGS) technology has reinforced the development of the molecular genetic research in sesame in the recent two decades. In the early stage, most sesame scientists endeavored to develop more specific DNA markers with high polymorphism and to construct the molecular genetic maps for sesame with the aid of the limited expressed sequence tag (EST) and transcriptome information. In the recent decade, a lot of sesame genome information stimulated rapid progresses in the molecular genetics analysis of important agronomic traits in sesame. In this chapter, we overview the main achievements of the high-resolution molecular genetic map construction, quantitative trait loci (QTLs) mapping related to the key traits, map location of molecular markers, and identification of the key functional genes in sesame. Especially, several successful examples for gene location and detection of molecular markers associated with yield, seed quality, disease resistance, and other key agronomic traits in sesame are enumerated. The prospects and strategies for establishing the molecular breeding techniques in sesame are also discussed.

8.1 Introduction

Sesame is regarded as an orphan crop species in plant genome research because of the ancient phylogenetic position and the remote phylogenetic distance from cereals and other economic crops. In particular, sesame is a self-pollinated crop and has a relatively narrow genetic base (Yue et al. 2012; Zhang et al. 2012a, c; Uncu et al. 2015). Thus, to rapidly aggregate the elite genic alleles into a new variety using the conventional breeding techniques with the limited breeding materials seems difficult. In order to increase the breeding efficiency, digging amount of functional genes and molecular markers or QTLs significantly associated with the target traits is necessary. Therefore, clarifying the molecular regulation mechanism of the key agronomic traits and determining the theoretical basis of the molecular breeding techniques become more urgent in sesame.

H. Zhang (✉) · L. Wei · Y. Duan · H. Miao
Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, People's Republic of China
e-mail: zhanghaiyang@zzu.edu.cn

L. Wei
e-mail: libinwei2013@aliyun.com

Y. Duan
e-mail: duanyly@aliyun.com

Y. Zhao · T. Zhou
Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan, People's Republic of China
e-mail: zhaoyz63@163.com

H. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-98098-0_8

In recent years, the high-throughput next-generation sequencing (NGS) technology facilitated the progress in sesame molecular genetic research specifically including constructing high-resolution molecular genetic maps, locating the QTLs, detecting molecular markers, and identifying the key functional genes related to agronomic traits (Ganal et al. 2011; Wei et al. 2011; Sim et al. 2012; Wang et al. 2015; Zhao et al. 2015; Zhang et al. 2016). The accomplishment of the Sesame Genome Project further supplied the key genome information for genetic and genomic research and marker-assisted breeding in sesame (Zhang et al. 2012b, 2019; Miao 2014; Wang et al. 2014; Wei et al. 2015a, b). We here review the history of molecular marker development and the main achievement of molecular genetic analysis in sesame. The successful experiences provide a great deal of genomic information and the foundation for strengthening the future molecular breeding in sesame.

8.2 Molecular Marker Discovery in Sesame

Sesame is an underutilized and disparaged crop. Compared to cereals and other major oilseed crops, the investment in sesame breeding research and industry in the world is evidently meager. Till 2009, only 3328 sets of expressed sequence tag (EST) data of sesame were deposited in the NCBI dataset. Limited achievement of the genetics and genomics research in turn hinders the development of sesame breeding and improvement in production in the world. Before the initiation of the Sesame Genome Project (the SGP) in 2010, the molecular genetics and genomics research in sesame was lagging. Similar to other crops, detecting DNA markers with high polymorphism for genetic map construction and mapping of quantitative trait loci (QTLs) related to various agronomic traits are the core points during the early development of sesame molecular genetics and breeding.

8.2.1 Universal Molecular Markers

Before the development of the NGS technique, scientists evaluated the genetic diversity of sesame germplasm using phenotype and isozymes. In 1990s, the isozymes of isocitrate dehydrogenase (IDH) were successfully utilized for sesame germplasm evaluation. Two alleles of IDH were used in the assessment of the 68 accessions (Isshiki and Umezaki 1997). In the early stage of the molecular genetics research in sesame, the universal molecular markers, including the random amplified polymorphic DNA (RAPD), sequence characterized amplified region (SCAR), inter-simple sequence repeats (ISSR), amplified fragment length polymorphism (AFLP), and sequence-related amplified polymorphisms (SRAP) markers were gradually used for the genetic diversity analysis of sesame germplasm accessions (Bhat et al. 1999; Kim et al. 2002; Ercan et al. 2004; Hernan and Petr 2006). Bhat et al. (1999) compared the 58 sesame accessions using RAPD markers and found that the extent of genetic diversity of the Indian varieties was greater than that of the exotics. Kim et al. (2002) used 40 ISSR primers to determine the genetic relationship among 75 sesame accessions. Only 33% primers exhibited polymorphism, and the population structure reflected the low diversity and the narrow genetic basis of the 75 Korean accessions (Kim et al. 2002). Meanwhile, Zhang et al. (2010) evaluated a sesame core collection containing 404 varieties using the 11 SRAP and three simple sequence repeat (SSR) markers. The results indicated that 72% fragments were polymorphic, and the 404 varieties were grouped into seven groups. The Nei's gene diversity (h) varied from 0.1772 to 0.2613. Considering the polymorphism of the universal molecular markers in samples is unstable, some scientists further dig the species-specific DNA markers, such as simple sequence repeat (SSR) markers, single nucleotide polymorphism (SNP), and insertion/deletion (InDel) markers as substitutes for genetics research in sesame by the beginning of the twenty-first century.

8.2.2 SSR Marker

Of the species-specific markers, SSRs and EST-SSRs are the most powerful PCR-based DNA markers in plant diversity analysis. SSRs are short tandem repeats with 1–6 bp in length and have high abundance, variability, and reproducibility with the codominant nature in plant genomes (Powell et al. 1996; Wei et al. 2008; Zhang et al. 2012b). In sesame, the first report about the development of genomic SSRs for sesame was published by Dixit et al. (2005). About 20% (10) SSR markers developed from a sesame cDNA library showed polymorphic in the 16 accessions (Dixit et al. 2005). To dig more polymorphic SSR markers for sesame, Wei et al. (2008) analyzed all the 3,328 sesame EST sequences in NCBI database and developed 50 EST-SSR markers. The PCR results showed that only 27 (61.4%) EST-SSR markers were polymorphic in the 34 cultivated sesame accessions and two wild accessions. In 2012, the transcriptomes of 24 sesame samples were sequenced using RNA-seq techniques on Ilumina sequencing platform (Zhang et al. 2012b). The 42,566 uni-transcript sequences covering 47,987 Kb size in sesame genome were assembled and analyzed. As a result, a total of 7324 ($\geq$ 15 bp) and 4440 ($\geq$ 18 bp) SSRs were presented in the uni-transcripts (Zhang et al. 2012b). However, only 32 (11.59%) genic-SSR markers were polymorphic in 24 cultivars. The polymorphism level of SSRs in sesame was similar to those of wheat and peanut (Gupta et al. 2010; Liang et al. 2020).

In the initiating stage, in order to develop more DNA markers with high polymorphism, many scientists employed an enrichment strategy to develop the microsatellite makers and SSR primers (Spandana et al. 2012; Badri et al. 2014; Surapaneni et al. 2014). Spandana et al. (2012) used multi-enzymes digestion, microsatellite oligo probes, and streptavidin magnetic beads to enrich microsatellite isolation in sesame. More than 200 SSR motifs ($\geq$ 2 repeat units or 6 bp) were detected. Surapaneni et al. (2014) constructed three sesame SSR-enriched libraries, i.e., (AG)n, (CT)n, and (AGC)n using the selective hybridization method. About 75 clones contained the microsatellite repeats, and 62 clones were found having unique flanking regions and selected for primer design.

With the wide application of NGS technique in sesame, more and more cDNA libraries were sequenced. As a great amount of transcriptomes and genome data of sesame have been deposited in the public dataset, the development and application of the specific molecular markers in sesame diversity analysis are significantly improved (Ke et al. 2011; Wei et al. 2011; Wang et al. 2012; Zhang et al. 2012b; Yepuri et al. 2013; Wei et al. 2014a; Wu et al. 2014a; Uncu et al. 2015). In particular, Ke et al. (2011) obtained 32,421 uniESTs with the cDNA library of immature sesame seeds and identified 1949 non-redundant (NR) SSRs in 1688 uniESTs, of which 349 primer pairs were designed for genetic diversity analysis in sesame (Wu et al. 2014a). Results indicated that 11 (3.5%) of 349 primer pairs were found polymorphic in 130 sesame accessions. Meanwhile, Wei et al. (2011) sequenced the transcriptomes of the mixture samples including root, leaf, shoot tip, flower, and the developing seed using Illumina paired-end sequencing technology. A total of 7702 SSRs were identified from 86,222 consensus sequences. Of the 50 primer pairs, 40 microsatellite loci showed allelic polymorphism.

Moreover, based on the genome information (var. Zhongzhi no. 13, WGS raw reads deposited no. SRA122008), 23,438 SSRs ($\geq$ 5 repeats) were identified (Wang et al. 2014; Wei et al. 2014b). In sesame genome, excluded the mononucleotide repeats and complex SSR types, dinucleotide was the most common repeat motif with the high frequency of 84.24%. Followed was trinucleotide with the frequency of 13.53%. The evaluation results showed that 218 (14.53%) of the 1,500 designed SSR primer pairs were polymorphic in the four accessions. Subsequently, Wang et al. (2017) updated the genome data and re-analyzed the genomic SSR loci in sesame genome. A total of 110,495 SSRs were detected. Mononucleotide (39.1%) and dinucleotide (34.3%) repeats were the most common repeats. Meanwhile, Uncu et al. (2015) sequenced the genome of 'MMuganli 57' using

Roche 454 GS-FLX and identified 19,816 SSRs from the 65 Mb genome data. The mononucleotide repeats were the most abundant, representing 48.5% of all SSRs. Of the designed 849 primer pairs, 228 (26.9%) were polymorphic between *S. indicum* and *S. mulayanum* (Uncu et al. 2015). To favor for the application of the polymorphic SSRs, Dossa et al. (2017) established an online database SisatBase (http://www.sesame-bioinfo.org/SisatBase/) with thousands of polymorphic SSRs. All the above SSRs and validated polymorphic SSRs provide useful information for genetic diversity assessment and molecular-assisted breeding (MAS) in sesame.

8.2.3 SNP and InDel Marker

SNP and InDel markers indicate the abundant base variants in genomes and are being utilized for genetics and genomics analyses in organisms (Rafalski 2002; Riahi et al. 2013). Accomplishment of the reference genome stimulated the aggregation of the genomics data in sesame. Based on the transcriptome and genome data, a great amount of SNP and InDel markers were dug and applied for genetic map construction and the genetic analysis of key traits in sesame (Wei et al. 2014a; Wu et al. 2014a; Zhang et al. 2016; Du et al. 2019). Wei et al. (2014a) constructed a reference transcriptome by the de novo assembly technique using the published three RNA-Seq datasets in sesame. About 10,950 SNPs in 4660 contigs and 590 InDels in 524 contigs presented in the alignments between 'Rongxian black sesame' (RXBS) and 'Zhongzhi 11'. Meanwhile, the 40 SNP primer sets were designed and applied for PCR screening in sesame for the first time. As a result, 24 (60.0%) SNP primer sets appeared allelic in three accessions, while 17 (85%) InDel (InDel length > 3 bp) primer pairs showed the evident polymorphism (Wei et al. 2014a).

For sesame, genome resequencing results indicated that up to millions of SNPs existed among the cultivated sesame varieties (Wei et al. 2015a, b; Zhang et al. 2016). The chromosome-scaled genome map of sesame (var. Yuzhi 11) and the genome resequencing data of sesame accessions improved the SNP and InDel application for gene cloning using the tested population and genome evolution analysis in *Sesamum* (Zhang et al. 2016; Zhang et al. 2018, 2019; Miao et al. 2019).

8.3 Molecular Genetic Linkage Maps in Sesame

Constructing a genetic linkage map is a prerequisite for exploring the trait inheritance at a genome-wide level (Sim et al. 2012; Verma et al. 2015). A molecular genetic map, especially a high-density molecular genetic map, provides the important genetic inheritance information in the tested population (Zhang et al. 2013b). Fine mapping QTLs or candidate genes related to the quantitative traits through high-resolution genetic maps are still an efficient method for genetic inheritance analysis of key quantitative traits in plants (Sim et al. 2012). Moreover, the application of NGS technology improves molecular marker digging and high-density SNP genetic maps construction in crops (Ganal et al. 2011; Sim et al. 2012; Wang et al. 2015; Zhao et al. 2015). For sesame, with the continuous development of the DNA marker detection techniques, more than ten molecular genetic maps have been constructed using EST-SSR, AFLP, RSAMP, SSR, or SNP markers till now.

8.3.1 SSR Genetic Maps

For sesame, the first molecular genetic linkage map for sesame was constructed using an F_2 population derived from the cross 'COI1134' 'RXBS' in 2009 (Wei et al. 2009). A total of 8 EST-SSRs, 25 AFLP markers, and 187 RSAMPLs (random selective amplification of microsatellite polymorphic loci) markers were located in the 30 linkage groups (LGs). In 2012, the genetic map was improved with 260 F_2 individuals (Zhang et al. 2013b). A total of 653 markers including 30 EST-SSR, 50 AFLP, and 573 RSAMPL markers were located on 14 LGs

with a total length of 1216.00 cM. Meanwhile, four QTLs linked to the seed coat color were mapped to three linkage groups (i.e., LG1, LG11, and LG13) for the first time (Zhang et al. 2013b).

Subsequently, another SSR genetic map was constructed using the sixth generations of a recombinant inbred line (RIL) population between 'Zhongzhi no. 13' and 'Yiyangbai' in sesame (Zhang et al. 2014). Of these 70 polymorphic SSRs, SRAP markers and AFLP markers were distributed into the 15 LGs. Based on the genetic map, six QTLs related to waterlogging tolerance at flowering stage were located on four LGs (LG 7, 9, 13, and 15). The phenotypic variance explained of the six QTLs varied from 5.67 to 17.19%. Furthermore, Wang et al. (2017) used thousands of the developed SSR markers and constructed the third SSR genetic map using 548 RIL lines of a cross between 'ZZM2748' and 'Zhongzhi no. 13'. The results showed that only 6.77% SSRs were polymorphic between the two parents and indicated the relatively low diversity of sesame accessions. In the genetic map, 424 polymorphic SSR markers were located on the 13 LGs, with a total length of 1869.8 cM. In addition, 14 QTLs associated with the resistance to charcoal rot disease were detected, with the contribution rate ranging from 3 to 14.16% in four field environments (Wang et al. 2017).

8.3.2 High-Density SNP Genetic Maps

In the past a few years, the reduced representation genome sequencing (RRGS) flourished and is being applied for SNP genetic map construction in sesame (Zhang et al. 2013b; Wu et al. 2014b; Uncu et al. 2016; Wang et al. 2016a; Mei et al. 2017; Du et al. 2019). At present, seven high-density SNP genetic linkage maps have been constructed and reported for sesame using the simplified genome sequencing techniques (Table 8.1). The first SNP genetic map for sesame was constructed in 2013 (Zhang et al. 2013b). A total of 107 progeny of an F_2 population derived from the cross between 'Zhongzhi no. 13' and 'Shang dong Jiaxiang sesame' were sequenced using the specific length amplified fragment sequencing (SLAF-seq) platform (hereinafter named 'SLAF map'). A total of 1,079 SNPs were mapped onto 15 LGs, with a total length of 1474.87 cm. At the same time, another SNP map was constructed using restriction site-associated DNA sequencing (RAD-seq) platform (herein named 'RAD map') with 224 individuals of an RIL population derived from the cross 'Zhongzhi 14' × 'Miaoqianzhima'. The RAD map contained 1,230 markers in 14 LGs with the length of 844.46 cm (Wu et al. 2014b). Based on the RAD map, 13 QTLs for the seven yield-related traits were determined using multiple interval mapping (MIM) and the mixed linear composite interval mapping (MCIM) methods, respectively. Of the 13 QTLs, three major QTLs presented the high explanation rate ($R^2 > 10.0\%$ in MIM method and $ha^2 > 5.0\%$ in MCIM). Moreover, two co-localized QTL groups were identified and partially explained the correlation relationship among the five yield-related traits (Wu et al. 2014b).

Subsequently, four more genetic maps are constructed using RAD-seq, SLAF-seq, or GBS (genotyping by sequencing) approaches (Uncu et al. 2016; Wang et al. 2016a; Mei et al. 2017; Du et al. 2019). Up to nine thousand SNP markers were determined on 13 LGs of the SNP genetic maps (Mei et al. 2017). Based on the SNP genetic maps, QTL mappings of some key agronomic traits such as yield and yield components, seed quality, disease resistance, stress tolerance, and other traits have been performed in sesame (Table 8.1). However, all the above SNP genetics maps are unsaturated with low marker density.

8.3.3 Ultra-Dense SNP Genetic Map

In order to construct an ultra-dense SNP genetic map, Zhang et al. (2016) resequenced the genomes of 'Yuzhi DS899' and 'JS012' and their 120 progeny of the F_2 population using Illumina sequencing platform (Table 8.1). The draft genome map of var. Yuzhi 11 (PRJNA315784, version 2.0) was applied as the genome reference for SNP genetic map construction. Genome

Table 8.1 Overview of SNP genetic maps constructed for sesame

Population	Sample size	Sequencing method	Linkage group number	Marker number	Total length (cm)	QTL-related trait	References
F_2	107	SLAF-seq	15	1233	1474.87	/	Zhang et al. (2013b), Wang et al. (2016)
RIL	224	RAD-seq	14	1230	844.46	Seed yield and quality-related traits	Wu et al. (2014b, 2017)
RIL	430	RAD-seq	13	1522	1090.99	Plant height, seed coat color, and capsule size	Wang et al. (2016), Yang et al. (2017)
RIL	91	GBS	13	432	914	/	Uncu et al. (2016)
BC_1	150	SLAF-seq	13	9378	1974.23	Flower number per axil and branching habit	Mei et al. (2017)
F_2	122	SLAF-seq	13	159	2128.51	17 traits related to seed coat color, seed size, and 100 seed thousand	Du et al. (2019)
F_2	120	Whole-genome sequencing	13	30,193	2981.28	Determinate growth habit	Zhang et al. (2016)

comparison indicated that a total of 192,744 variants (including SNPs and InDels) were detected between 'Yuzhi DS899' and reference genome (var. Yuzhi 11), while 781,528 SNPs and InDels were found between the other parent 'JS012' and var. Yuzhi 11 for the relative distant relationship. As a result, the first ultra-dense SNP map for sesame comprised 3,041 bins (30,193 SNPs) on the 13 LGs, with a total length of 2981.28 cm (Table 8.1) (Fig. 8.1). The average marker density was 0.10 cm per SNP or 0.98 cm per bin marker. The SNP genetic map has been applied for genome assembly and integration with the cytogenetic map in sesame (described in the following chapters). Meanwhile, the ultra-dense SNP has been used to locate a *Sidt1* gene regulating the inflorescence determinacy in sesame and exhibits the effective map-based gene cloning method in sesame (Zhang et al. 2016).

Recently, Liang et al. (2020) published another ultra-dense SNP map using the two parents and 180 progeny of an RIL population through the whole-genome resequencing platform. A total of 1354 bins covering 538,090 variants were located onto 13 chromosomes. The density of bin markers was 0.98 cm/locus, covering an average physical length of 158.74 Kb per bin. Meanwhile, the genetic map was applied for locating the QTLs associated with PEG-induced drought tolerance-related traits during early seedling stage in sesame (Table 8.2).

8.4 QTL Mapping of Agronomic Traits in Sesame

In order to explore the genetic inheritance of agronomic traits in sesame, sesame scientists initially applied the genetic mapping method to determine the QTLs of the agronomic traits (Table 8.2). Till now, a total of 138 QTLs with the high phenotypic variance explanation ($R^2 \geq 5\%$) associated with the 44 agronomic traits in sesame have been reported (Table 8.2).

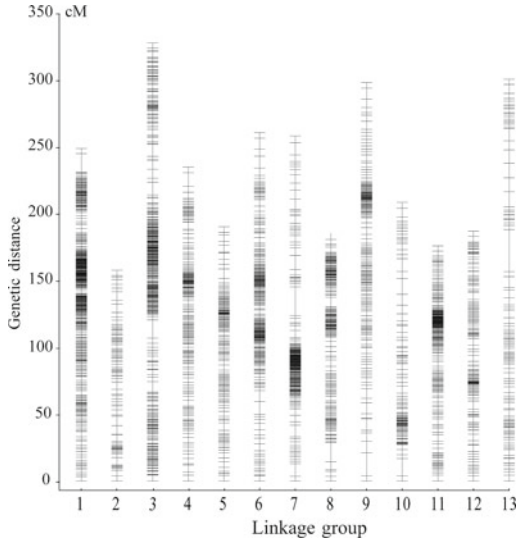


Fig. 8.1 Distribution of SNP markers in the 13 linkage groups of the ultra-dense SNP map for sesame. Bars indicate SNP markers. X-axis indicates the linkage group (LG). Y-axis indicates the genetic distance in centi Morgans (cm). A total of 30,193 SNP markers were mapped to 13 LGs. The 13 LGs were randomly designated as LG1 ~ LG13. (Cited from Zhang et al. (2016))

8.4.1 QTLs Location of Yield-Related Traits

As to the yield-related traits, Wu et al. (2014b) investigated the seven yield components traits including plant height (PH), height to the first capsule (HFC), capsule axis length, capsule number per plant (CNP), thousand seed weight (TSW), seed number per capsule (SNC), and capsule length under five environments during 2012–2013. Thirteen QTLs were detected on ten LGs for the above yield-related traits using MIM and MCIM methods. Of the 13 QTLs, three major QTLs exhibited high phenotypic variance with the high $R^2 > 10.0\%$ in MIM and $R^2 > 5.0\%$ in MCIM, respectively. Moreover, two co-localized QTL groups were found for the five yield-related traits (Wu et al. 2014b). For PH, two QTLs (Qph-6, Qph-12) with R^2 ranging from 5.6 to 9.1% using MIM were located in LG 6 and LG 12. Meanwhile, Wang et al. (2016) identified 41 QTLs for PH trait under three environments based on the RAD map with the RIL sesame population. Of the nine QTLs with the R^2 ranging from 5 to 24%, two loci qPH-8.2 (in LG 8) and qPH-3.3 (in LG 3) were the main QTLs and showed 23% and 18% of the trait variation, respectively. Both results reflected that the inheritance of the PH trait was controlled by several major genes and multiple minor effect genes with additive effects in sesame.

Furthermore, using the RIL population of 'Zhaongzhi no. 13' and 'ZZM2748' and the SSR genetic map, Wang et al. (2017) and Yang et al. (2017) analyzed the inheritance of the three capsule size-related traits for the central and lateral capsules in plants, respectively, under three environments. A total of 15 QTLs with high R^2 values of more than 5% contributed to the six group traits. A locus (namely qMCL9.1, qFCL9.1 and qFCW9.1, and qFCT9.1 for the corresponding traits) in interval SLG_bin 105-bin 106 was linked to the length of central capsules and the length, width, and thickness of lateral capsules in all three environments (Yang et al. 2017).

Interestingly, Mei et al. (2017) studied the branching habit and flower number per leaf axil using the constructed 150 BC_1 population. In the high-resolution SNP genetic map through SLAF-seq, two co-segregated SNP markers linked with both traits were detected (Mei et al. 2017).

8.4.2 QTLs Location of Seed Quality Trait

For sesame seed, the oil content, protein content, and sesamin content are the attractive traits for high-quality variety breeding. Wu et al. (2017) detected eight QTLs for the above three seed quality traits with heritability of additive effects ranging from 0.41 to 14.55% using MCIM method. In MIM method, 13 QTLs were mapped on nine LGs. The explanation ratio varied from 5.2 to 18.6% of the total phenotypic variation (Wu et al. 2017).

The seed coat color trait is another important seed quality trait in sesame (Zhang et al. 2012a,

Table 8.2 Information of QTL (gene) mapping of agronomic traits in sesame

Trait type	Trait	QTL locus	LG position	Marker interval (QTL region, cm)	Explanation ($\geq 5\%$) or P-value	References
Yield-related traits	Thousand seed weight	Qtgw-11	LG11	SBN1798-SBN1765 (17.9–21.2)	7.7–12.3	Wu et al. (2014b)
		qtsw4	LG04	MK1268296-MK1268983 (57.19–57.69)	15.09	Du et al. (2019)
		qtsw9	LG09	MK193210- MK167922 (120.45–120.86)	6.90	
		qtsw12	LG12	MK1695007-MK1754691 (124.45–124.92)	19.56	
	Grain number per capsule	Qgn-6	LG6	SBN1261-SBI043 (74.4–99.0)	80.-18.3	Wu et al. (2014b)
	Plant height	qPH-3.1	SLG3	SLG3_bin98-SLG3_bin99 (80.91)	6	Wang et al. (2016a)
		qPH-3.2	SLG3	SLG3_bin114-SLG3_bin115 (92.41)	15	
		qPH-3.3	SLG3	SLG3_bin126-SLG3_bin127 (98.01)	18	
		qPH-4.1	SLG4	SLG4_bin49-SLG4_bin50 (34.11)	8	
		qPH-4.2	SLG4	SLG4_bin56-SLG4_bin57 (42.61)	6	
		qPH-8.1	SLG8	SLG8_bin107-SLG8_bin108 (65.61)	9	
		qPH-8.2	SLG8	SLG8_bin111-SLG8_bin112 (70.51)	23	
		qPH-9.1	SLG9	SLG9_bin102-SLG9_bin103 (78.81)	9	
		qPH-9.2	SLG9	SLG9_bin105-SLG9_bin106 (84.71)	10	
	Capsule zone length	qCZL-3.1	SLG3	SLG3_bin126-SLG3_bin127 (99.00)	9	
		qCZL-4.1	SLG4	SLG4_bin48-SLG4_bin49 (34.00)	7	
		qCZL-8.1	SLG8	SLG8_bin111-SLG8_bin112 (71.50)	24	
		qCZL-9.1	SLG9	SLG9_bin103-SLG9_bin104 (79.60)	6	
		qCZL-13.1	SLG13	SLG13_bin32-SLG13_bin33 (24.40)	7	
	Height of the first capsule node	qHFC-3.1	SLG3	SLG3_bin112-SLG3_bin113 (92.21)	9	
		qHFC-3.2	SLG3	SLG3_bin126-SLG3_bin127 (97.01)	22	

(continued)

Table 8.2 (continued)

Trait type	Trait	QTL locus	LG position	Marker interval (QTL region, cm)	Explanation ($\geq 5\%$) or P-value	References
		qHFC-4.1	SLG4	SLG4_bin45-SLG4_bin46 (32.41)	5	
		qHFC-8.1	SLG8	SLG8_bin107-SLG8_bin108 (65.71)	5	
		qHFC-8.2	SLG8	SLG8_bin111-SLG8_bin112 (69.51)	9	
		qHFC-9.1	SLG9	SLG9_bin102-SLG9_bin103 (78.81)	10	
		qHFC-9.2	SLG9	SLG9_bin105-SLG9_bin106 (84.71)	11	
		qHFC-9.3	SLG9	SLG9_bin112-SLG9_bin113 (91.01)	10	
	Internode length	qIL-3.1	SLG3	SLG3_bin53-SLG3_bin54 (41.10)	7	
		qIL-8.1	SLG8	SLG8_bin111-SLG8_bin112 (71.50)	9	
		qIL-9.1	SLG9	SLG9_bin105-SLG9_bin106 (83.70)	10	
	Node number	qNN-3.1	SLG3	SLG3_bin111-SLG3_bin112 (91.80)	5	
		qNN-3.2	SLG3	SLG3_bin126-SLG3_bin127 (100.10)	10	
		qNN-8.1	SLG8	SLG8_bin110-SLG8_bin111 (66.50)	7	
		qNN-13.1	SLG13	SLG13_bin32-SLG13_bin33 (24.40)	9	
	Tip length without the capsule	qTL-3.1	SLG3	SLG3_bin124-SLG3_bin125 (94.40)	8	
		qTL-3.2	SLG3	SLG3_bin131-SLG3_bin132 (103.60)	8	
	Capsule length	Qcl-12	LG12	ZM1466-SBI005 (14.0–18.0)	52.2–75.6	Wu et al. (2014b)
	Length of the central capsule	q MCL9.1	SLG9	SLG9_bin105- bin106 (82.7)	7.74	Yang et al. (2017b)
		qMCL10.1	SLG10	SLG10_bin23- bin24 (19.3)	5.19	
		qMCL12.1	SLG12	SLG12_bin74- bin75 (54.4)	8.43	
		qMCL13.1	SLG13	SLG13_bin78- bin79 (50.3)	13.69	
	Width of the central capsule	qMCW9.1	SLG9	SLG9_bin104- bin105 (81.1)	13.77	
		qMCW13.1	SLG13	SLG13_bin80- bin81 (50.8)	12.26	

(continued)

Table 8.2 (continued)

Trait type	Trait	QTL locus	LG position	Marker interval (QTL region, cm)	Explanation ($\geq 5\%$) or P-value	References
	Thickness of the central capsule	q MCT9.1	SLG9	SLG9_bin112-bin113 (90.8)	8.44	
		q MCT13.1	SLG13	SLG13_bin80- bin81 (50.8)	8.93	
	Length of the lateral capsule	q FCL3.1	SLG3	SLG3_bin103- bin104 (84.5)	5.48	
		q FCL11.1	SLG11	SLG11_bin105- bin106 (69.6)	9.31	
	Width of the lateral capsule	q FCW3.1	SLG3	SLG3_bin110-bin111 (91.5)	7.05	
		q FCW9.1	SLG9	SLG9_bin105-bin106 (83.7)	19.65	
		q FCW13.1	SLG13	SLG13_bin78- bin79 (50.3)	8.48	
	Thickness of the lateral capsule	q FCT9.1	SLG9	SLG9_bin105- bin106 (83.7)	9.41	
		q FCT10.1	SLG10	SLG10_bin22- bin23 (18.2)	5.29	
	Determinate growth habit	*SiDt*	LG8	GenBank accession no. KU240042	100	Zhang et al. (2016)
	Branching habit	qBH-LG5	LG5	Marker41538–Marker31462 (1.75 - 2.75)	78.64	Mei et al. (2017)
	Flower number per leaf axil	SiFA	LG11	Marker34507 (co-segregated)	100	
Seed quality	Oil content	Qoc-1	LG1	SBN2389–SBN2485 (37.0–40.6)	10.1	Wu et al. (2017)
		Qoc-2	LG2	SBN2776–SBN1045 (13.5–15.2)	6.5	
		Qoc-5	LG5	SBN3585–SBN1490 (41.5–52.1)	7.7	
		Qoc-9	LG9	SBN1388–SBN1459 (6.9–9.7)	8.6	
		Qoc-16	LG16	SBN1927–SBN3232 (22.6–26.6)	7.6	
	Protein content	Qpc-1	LG1	SBN741–SBN2389 (20.0–28.7)	7.6	
		Qpc-2	LG2	SBN2776–SBN2749 (13.5–15.2)	6.2	
		Qpc-3	LG3	HS184–SBI013 (6.3–9.7)	5.7	
		Qpc-5	LG5	SBI007–SBI057 (62.8–68.4)	10.9	
		Qpc-6	LG6	SBN3210–SBN636 (94.0–102.4)	6.4	

(continued)

Table 8.2 (continued)

Trait type	Trait	QTL locus	LG position	Marker interval (QTL region, cm)	Explanation ($\geq 5\%$) or P-value	References
	Sesamin content	*Qsc-4*	LG4	SBI050–SBN1100 (70.7–73.7)	5.4	
		Qsc-5	LG5	ZHY01–SBN1548 (49.4–56.2)	11.1	
		Qsc-5	LG5	SBN3568–SBI007 (40.5–62.9)	18.6	
		Qsc-5	LG5	SBN3568–SBI007 (40.5–62.9)	14.1	
		Qsc-8	LG8	SBN1735–SBN2668 (35.4–44.6)	8.9	
		Qsc-8	LG8	SBN1407–SBN1140 (37.0–38.2)	5.2	
Seed coat color	RGB value	QTL1-1	LG1	Y1991F/R - Hs1015F/E9 (139.6–141.9)	23.32–39.95	Zhang et al. (2013b)
		QTL11-1	LG11	Hs1125R/E11 - Y2017F/M11 (18.6–30.8)	9.72–20.61	
		QTL11-2	LG11	Y2017F/M11 - Y640/ E16 (32.2–51.2)	9.6–31.86	
		QTL13-1	LG13	Hs1097F/E1 - Y2632F/M7 (34.7–43.7)	12.8–30.56	
	a* color value	qSCa-4.1	SLG4	SLG4_bin63-SLG4_bin64 (50.90)	13	Wang et al. (2016a)
		qSCa-8.1	SLG8	SLG8_bin114-SLG8_bin115 (73.40)	25	
		qSCa-8.2	SLG8	SLG8_bin105-SLG8_bin106 (62.60)	9	
	b* color value	qSCb-4.1	SLG4	SLG4_bin63-SLG4_bin64 (50.90)	39	
		qSCb-8.1	SLG8	SLG8_bin114-SLG8_bin115 (72.40)	21	
	L* color value	qSCl-4.1	SLG4	SLG4_bin63-SLG4_bin64 (50.90)	21	
		qSCl-8.1	SLG8	SLG8_bin114-SLG8_bin115 (73.40)	46	
		qSCl-11.1	SLG11	SLG11_bin1-SLG11_bin2 (0.00)	14	
	a* color value	qscca*4	LG04	MK1269895-MK1382037 (79.58–79.99)	19.09	Du et al. (2019)
		qscca*9	LG09	MK1792520- MK290567 (80.47–90.56)	20.02	
	b* color value	qsccb*4	LG04	MK1281005-MK1338566 (78.19–78.68)	23.10	

(continued)

Table 8.2 (continued)

Trait type	Trait	QTL locus	LG position	Marker interval (QTL region, cm)	Explanation ($\geq 5\%$) or P-value	References
	L* color value	qsccL*4	LG04	MK1281005-MK1338566 (78.19–78.68)	8.56	
	a color value	qscca4-1	LG04	MK1303398-MK1353258 (80.89–81.38)	17.43	
	a color value	qscca4-2	LG04	MK1269895-MK1382037 (79.58–79.99)	19.79	
	b color value	qsccb4	LG04	MK1303398-MK1353258 (80.89–81.38)	17.65	
	L color value	qsccL4	LG04	MK1303398-MK1353258 (80.89–81.38)	7.62	
		qsccL4	LG04	MK1281005-MK1338566 (78.19–78.68)	7.48	
	X color value	qsccX4	LG04	MK1303398-MK1353258 (80.89–81.38)	8.27	
	Y color value	qsccY9	LG09	MK183845- MK253698 (101.72–103.68)	33.25	
		qsccY4	LG04	MK1303398-MK1353258 (80.89–81.38)	7.63	
	Z color value	qsccZ9	LG09	MK1791984- MK253852 (90.56–90.97)	32.88	
		qsccZ12	LG12	MK1580955-MK1696180 (159.98–161.95)	5.58	
Seed size	Seed area	qsa1	LG01	MK1149758-MK1110290 (36.27–41.19)	6.64	Du et al. (2019)
		qsa11	LG11	MK21445- MK39038 (165.41–166.23)	13.86	
	Seed circularity	qsc1	LG01	MK1130242-MK1199094 (183.42–183.92)	12.16	
		qsc2	LG02	MK111106- MK635618 (16.77–17.26)	12.11	
	Seed diameter	qsd5	LG05	MK714015- MK804273 (22.27–22.27)	6.43	
		qsd11	LG11	MK21445- MK39038 (165.41–166.23)	14.70	
	Seed length	qsl11	LG11		15.12	

(continued)

Table 8.2 (continued)

Trait type	Trait	QTL locus	LG position	Marker interval (QTL region, cm)	Explanation ($\geq 5\%$) or P-value	References
				MK21445- MK39038 (165.41–166.23)		
		qsl13	LG13	MK1860058-MK1843260 (66.41–66.90)	7.45	
	Seed length-to-width ratio	ql/w1	LG01	MK1394203-MK1213353 (172.85–173.34)	11.76	
		ql/w2	LG02	MK111106- MK635618 (16.77–17.26)	11.43	
	Seed perimeter	qsp11	LG11	MK84783- MK21445 (164.51–165.41)	13.97	
	Seed width	qsw1	LG01	MK1149758-MK1219420 (36.27–39.71)	6.73	
		qsw5	LG05	MK714015- MK804273 (22.27–22.27)	8.97	
		qsw11	LG11	MK21445- MK39038 (165.41–166.23)	12.39	
Disease resistance	Charcoal rot	*qCRR3.2*	LG3	ZMM5636 ~ ZMM5775 (39.30)	12	Wang et al. (2017)
		qCRR3.3	LG3	ZMM2218 ~ ZMM4682 (52.30)	10	
		qCRR3.4	LG3	ZMM4682 ~ ZMM5444 (58.40)	9	
		qCRR8.1	LG8	ZMM5060 ~ ZMM5061 (10.50)	5	
		qCRR8.2	LG8	ID0041 ~ ZM638 (115.70)	5	
		qCRR8.3	LG8	ZM638 ~ ZMM1682 (123.70)	5	
		qCRR9.1	LG9	ZMM2323 ~ ZMM0205 (104.70)	8	
		qCRR12.1	LG12	ID0046 ~ ID0133 (53.80)	6	
		qCRR12.2	LG12	ZMM0913 ~ ZMM3752 (89.80)	14	
		qCRR13.2	LG13	ZMM2344 ~ ZMM2343 (73.50)	8	
Stress tolerance	Waterlogging tolerance	qEZ09ZCL13	LG13	ZM22-ZM92 (0.0)	10.20	Wang et al. (2017)
		qWH09CHL15	LG15	E16M19-314M14a (8.0)	7.55	
		qEZ10ZCL07	LG7	E5M12a-ZM351 (4.5)	8.14	
		qWH10ZCL09	LG9	M20E10-ZM428 (7.0)	5.67	
		qEZ10CHL07	LG7	E5M12a-ZM351 (4.5)	6.69	

(continued)

Table 8.2 (continued)

Trait type	Trait	QTL locus	LG position	Marker interval (QTL region, cm)	Explanation ($\geq 5\%$) or P-value	References
		Qwh10CHL09	LG9	M20E10-ZM428 (7.0)	17.19	
Drought tolerance	Fresh seedling weight treated by PEG	qSWP1	Chr1	c01b003- c01b010	8.20	Liang et al. (2020)
		qSWP3	Chr3	c03b116- c03b119	6.61	
		qSWP9	Chr9	c09b031- c09b040	7.50	
	Shoot length treated by PEG	qSLP1	Chr1	c01b032- c01b035	8.26	
		qSLP8	Chr8	c08b055- c08b063	5.83–5.88	
		qSLP9-1	Chr9	c09b015- c09b021	7.26	
		qSLP9-2 c	Chr9	c09b031- c09b033	11.00	
	Root length treated by PEG	qRLP1	Chr1	c01b062- c01b070	6.77–7.97	
		qRLP6	Chr6	c06b054- c06b060	5.15	
		qRLP7	Chr7	c07b030- c07b036	6.40–6.61	
		qRLP12	Chr12	c12b032- c12b036	11.85–14.46	

Modified from Zhang et al. (2019)

2013a). Initially, Zhang et al. (2013a) evaluated the RGB values in the F_2 and two F_3 families and identified four stable QTLs for seed coat color in sesame. Of these QTLs, QTL 1–1 presented the highest explanation of the phenotypic variation (23.32–39.95%) followed by QTL 13-1 (12.8–30.56%) with CIM method. In CIM method, the explanation to the additive heritability of the four loci (i.e., QTL 1-1, QTL 11-1, QTL 11-2, and QTL 13-1) was 19.4–30.03%, 13.9–21.22%, 5.56–19%, and 9.13–21.02%, respectively. Subsequently, Wang et al. (2016) measured the L, a*, and b* values of a RIL population. Similar to the results of Zhang et al. (2013a), four loci linked with the L*, a*, and b* traits were detected on three LGs. The highest contribution rate of the loci was 46%.

8.4.3 QTL Location of Male Sterile (MS) Trait

As to the novel GMS line (W1098A) controlled by one dominant GMS gene (*Ms*), Liu et al. (2015) screened 13 SSR markers linked to the *ms* gene in near-isogenic lines W1098AB using BSA (bulked segregant analysis) technique. The target markers close to the flanking sides of *Ms* were SBM_{298} and GB_{50} with the respective genetic distance of 0.15 and 0.70 cm, respectively, in the linkage map. The marker information provides the genetics basis for male sterility application and further gene location in sesame.

8.4.4 QTL Location of Charcoal Rot Resistance Trait

Wang et al. (2017) investigated the charcoal rot disease occurrence in 430 RIL lines of the cross between 'Zhaongzhi no. 13' and 'ZZM2748' in four environments under natural conditions. Ten QTLs ($R^2 \geq 5\%$) linked to charcoal rot resistance were detected. The highest R^2 was 14% in qCRR12.2. Of the ten QTLs, two loci of qCRR8.2 ($R^2 = 5\%$) and qCRR8.3 ($R^2 = 5\%$) presented in all four environments; two loci (qCRR9.1 and qCRR12.2) presented in three environments and atmosphere conditions. The results indicated that the genetic inheritance of charcoal rot resistance in sesame was complicated, as the DI phenotypes were easily affected by the pathogen infection and pathogen–plant interaction.

8.4.5 QTLs Location of Waterlogging Tolerance Trait

Zhang et al. (2014) detected a total of six QTLs associated with waterlogging tolerance at flowering stage using 260 lines of an RIL population from 'Zhaongzhi no. 13' and 'Yiyangbai'. The six QTLs were located on LG 7, 9, 13, and 15 with the phenotypic variance ranging from 5.67 to 17.19%. Of which QTL qWH10CHL09 explained the highest phenotype variance. Moreover, SSR marker ZM428 was detected and closely linked to qWH10CHL09 with a genetic distance of about 0.7 cM.

8.4.6 Gene Cloning Using SNP Genetic Maps

Besides guiding the genome assembly, the first ultra-dense SNP genetic map has been applied for cloning the *SiDt* gene regulating the inflorescence determinacy trait in sesame (Zhang et al. 2016). Association results indicated that the locus was located in the inheritance interval of 16.7–22.2 cm in LG8. The R^2 value reached 70.2%. Meanwhile, QTL network results also determined a locus in the 18.0–19.2 cM inheritance interval of LG8 with the *P* value of $\leq$ 1E-6. The determination of the inheritance interval and the physical distance of the QDt1 provided the basis for *SiDt* gene cloning. With the aid of screening population and gene alleles test, the target gene *SiDt* (GenBank accession no. KU240042) was finally determined (Zhang et al. 2016). *SiDt* gene encodes a terminal flower (TFL) like protein. Interestingly, the alleles of *SiDt* in the three genotypes with various inflorescence growth habits showed the different gene structures (Fig. 8.2).

The *SiDt* gene is 1809 bp and comprised of four exons and three introns (Fig. 8.2). In the determinate parent (Yuzhi DS899, *dt1*), a base change (G397A) happened in *Sidt1* gene. The SNP mutation finally resulted in change of the apical meristem development from indeterminacy to determinacy during late flowering stage (Fig. 8.2). As to *dt2* type (described in the Chap. 5) induced by gamma ray (5,000 gray), the *SiDt* gene and the flanking regions of 24.5 kb are missing (Fig. 8.2). The loss of *SiDt* gene resulted in the determinacy of the shoot apex during the early flowering stage (Melzer et al. 2008; Uzun and Çağırgan 2009; Zhang et al. 2016).

8.5 Association Mapping of Key Traits in Sesame

Association mapping, as well as the linkage analysis (QTL mapping) becomes the main analysis tool for dissecting complex phenotypic variation (Li et al. 2014). Differing from the traditional linkage mapping, association mapping approach can be used to precisely determine the molecular markers tightly associated with the

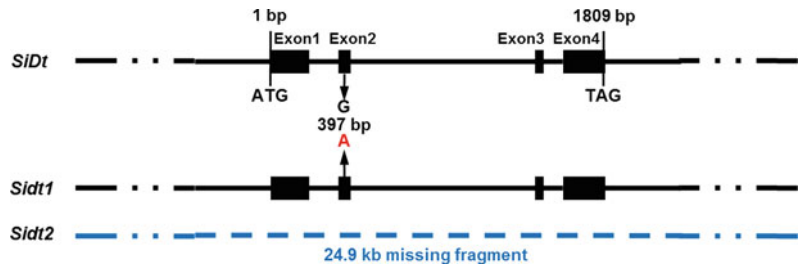


Fig. 8.2 Map-based cloning of *SiDt* and the alleles in sesame. *SiDt* (DS899s00170.023) gene is 1,809 bp and comprised of four exons (in black block) and three introns. For *Sidt1*, the G379A SNP site (in red) is located in the second exon. For *Sidt2* allele in *dt2* genotype, a 24.9 kb fragment including the *Sidt2* sequence (in blue broken line) in scaffold 00,170 is missing according to 08TP092 genomic data (PRJNA316751). (Modified from Zhang et al. (2016))

target traits using both natural population and mapping populations (Abdurakhmonov et al. 2009; Zhao et al. 2010; Miao et al. 2020). Till now, the association mapping studies are involved in seed quality, plant growth and development, and other key agronomic traits in sesame (Table 8.3).

Wei et al. (2012) initially performed the association analysis of the oil content within the 216 Chinese sesame accessions using 79 molecular primer pairs (including SSRs, SRAPs, and AFLPs) in 2008 and 2009. As a result, three SRAP and five SSR markers were detected associated with the oil content trait in two environments, with the high phenotypic variance of more than 5% ($P < 0.01$) (Wei et al. 2012). In order to systematically explore the genetic inheritance of oil content and protein content in seeds, Li et al. (2014) investigated the oil content and protein content traits in a total of 369 worldwide sesame germplasm accessions under five environments. A total of 112 polymorphic SSR markers were applied for the association mapping in the natural population. Comparison analysis results proved that the oil content is negatively correlated with the protein content. Similar to other crops, the oil and protein contents are controlled by major genes in sesame. Finally, 19 SSR markers were significantly associated with the oil content trait, with R^2 value ranging from 4 to 29%. Meanwhile, 24 markers were determined and associated with the protein content, with the R^2 ranging from 3 to 29%. Based on the reference sesame genome sequences, 17 SSR markers were neighboring to the loci of the plant lipid pathway genes, and the other two SSRs were located just next to a fatty acid elongation gene and a stearoyl-ACP desaturase gene, respectively. In the study, 36 candidate genes related to lipid pathway were screened for further analysis (Li et al. 2014).

Based on the universal markers and SSR markers, only small portion of the markers are found linked with the target traits under various environments (Zhang et al. 2012c; Liu et al. 2017). For instance, Zhang et al. (2012c) used the same population and markers described by Wei et al. (2012) to study the genetic inheritance of charcoal rot resistance. Only five–six markers were associated with charcoal rot disease index (DI) in 2011 using GLM (Q) and MLM (Q + K) model, respectively, with the high phenotypic variance of more than 5% (Zhang et al. 2012c). In addition, all the detected markers associated with drought tolerance at germination stage exhibited phenotypic variation ranging from 1.99% to 4.96% (Li et al. 2013). Liu et al. (2017) performed association mapping using 33 SSR markers and 18 drought-resistance phenotypes. A total of 120 loci and 63 loci were detected using GLM and MLM models, respectively. Only five markers significantly associated with the drought tolerance traits were detected by both the models, with the physical variance ranging from 4.07 to 13.66% (Liu et al. 2017).

Table 8.3 Association mapping analysis of agronomic traits in sesame

Population size	Marker type	Marker number	Traits	References
216	SSRs SRAPs AFLPs	79	Oil content; protein content; oleic acid concentration; linoleic acid concentration; charcoal rot resistance; drought tolerance	Zhang et al. (2012c), Wei et al. (2012, 2013), Li et al. (2013)
369	SSR	112	Oil content; protein contents	Li et al. (2014)
705	SNP	1,800,000	56 agronomic traits including yield, seed quality, and resistance traits	Wei et al. (2015a)
33	SSR	18	Drought resistance	Liu et al. (2017)
490	SNP	1,005,413	Salinity and drought tolerances	Li et al. (2018)
705	SNP	1,800,000	39 seed yield-related traits	Zhou et al. (2018)

Thus, Wei et al. (2015) performed the genome-wide association studies (GWAS) with the aid of sesame genome data and the genome resequencing techniques (Wang et al. 2014; Wei et al. 2015a, b; Zhou et al. 2018). For example, Wei et al. (2015a) detected a total of 549 associated loci for the 56 agronomic traits in 705 sesame accessions using the comprehensive GWAS method. Moreover, 46 candidate genes related to oil content, fatty acid biosynthesis, and yield traits were screened in sesame. Identification of lots of candidate genes for the key agronomic traits provides the important information for gene function analysis in sesame in the future.

8.6 Molecular Breeding Prospects

Precise selection of desirable and elite traits is very important for the plant breeders. However, most of the agronomic traits belong to quantitative traits and are controlled by polygenes with complex non-allelic quantitative effects. With the aid of NGS and the third-generation sequencing technologies, the genetics and genomics research in sesame has been improved in the recent a few years (Ganal et al. 2011; Sim et al. 2012; Zhang et al. 2013a, b; Miao 2014; Wang et al. 2014, 2015; Wei et al. 2015a, b; Zhao et al. 2015; Miao and Zhang 2016). Dossa (2017) gathered 151 published genomic sequence resources and QTLs and genes and compared the positions of the functional markers, QTLs, the candidate genes, and functional genes in 16 linkage groups in sesame. Finally, 15 clusters were determined (Dossa 2017). Especially, construction of the ultra-density SNP genetic maps (Zhang et al. 2016) and the achievement of the Sesame Genome Project pave the ways for QTL mapping and gene cloning of the key quantitative traits in sesame (Zhang et al. 2019). We believe that hundreds and thousands of genes and gene markers related to the important agronomic traits will be obtained and facilitated the molecular breeding in sesame (Zhang et al. 2013b; Wei et al. 2015a, b; Wang et al. 2016a; Zhang et al. 2016; Miao et al. 2020) in near future.

References

Abdurakhmonov IY, Saha S, Jenkins JN, Buriev ZT, Shermatov SE et al (2009) Linkage disequilibrium based association mapping of fiber quality traits in *G. hirsutum* L. variety germplasm. Genetica 136:401–417

Badri J, Yepuri V, Ghanta A (2014) Development of microsatellite markers in sesame (*Sesamum indicum* L.). Turk J Agri Forest 38(5):603–614

Bhat KV, Babrekar PP, Lakhanpaul S (1999) Study of genetic diversity in Indian and exotic sesame (*Sesamum indicum* L.) germplasm using random amplified polymorphic DNA (RAPD) markers. Euphytica 110:21–33

Dixit A, Jin MH, Chung JW, Jae-Woong YU, Chung HK et al. (2005) Development of polymorphic microsatellite markers in sesame (*Sesamum indicum,* L.). Mol Ecol Notes 5(4):736–738

Dossa K, Yu J, Liao B, Cisse N, Zhang X (2017) Development of highly informative genome-wide single sequence repeat markers for breeding applications in sesame and construction of a web resource: Sisatbase. Front Plant Sci 8(1470):1470

Du H , Zhang H, Wei L, Li C , Duan Y et al. (2019) A high-density genetic map constructed using specific length amplified fragment (SLAF) sequencing and QTL mapping of seed-related traits in sesame (*Sesamum indicum* L.). BMC Plant Biol 19(1):1–20

Ercan AG, Taskin M, Turgut K (2004) Analysis of genetic diversity in Turkish sesame (*Sesamum indicum* L.) populations using RAPD markers. Genet Resour Crop Evol 51(6):599–607

Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES et al. (2011) A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS ONE 6:e28334

Gupta PK, Kumar J, Mir RR, Kumar A (2010) Marker-assisted selection as a component of conventional plant breeding. Plant Breed Rev 33:145–217

Hernan EL, Petr K (2006) Genetic relationship and diversity in a sesame (*Sesamum indicum* L.) germplasm collection using amplified fragment length polymorphism (AFLP). BMC Genet 7:10

Isshiki S, Umezaki T (1997) Genetic variations of isozymes in cultivated sesame (*Sesamum indicum* L.). Euphytica 93(3):375–377

Ke T, Dong C, Mao H, Zhao Y, Chen H et al (2011) Analysis of expression sequence tags from a full-length-enriched cdna library of developing sesame seeds (*Sesamum indicum*). BMC Plant Biol 11(1):180

Kim DH, Zur G, Danin-Poleg Y, Lee S, Shim K et al (2002) Genetic relationships of sesame germplasm collection as revealed by inter-simple sequence repeats. Plant Breed 121:259–262

Li DH, Liu WP, Zhang YX, Wang LH, Wei WL et al (2013) Identification method of drought tolerance and

association mapping for sesame (*Sesamum indicum* L.). Acta Agron Sin 39(8):1425–1433

Li C, Miao HM, Wei LB, Zhang TD, Han XH et al (2014) Association mapping of seed oil and protein content in *Sesamum indicum* L using SSR markers. PLoS ONE 9 (8):e105757

Li DH, Dossa K, Zhang YX, Wei X, Wang L et al (2018) GWAS uncovers differential genetic bases for drought and salt tolerances in sesame at the germination stage. Genes 9(2):87

Liang J, Ye Y, Yan X, Yan T, Rao Y et al (2020) Construction of a high density genetic map and QTL mapping of PEG-induced drought tolerance at seedling stage in sesame using whole genome re-sequencing. Research Square. https://doi.org/10.21203/rs.3.rs-50720/v2

Liu H, Zhou X, Wu K, Yang M, Zhao Y (2015) Inheritance and molecular mapping of a novel dominant genic male-sterile gene in *Sesamum indicum* L. Mol Breed 35(1):9

Liu WP, Wei L, Li DH, Ren GX, Zhang YX et al (2017) Drought resistance of sesame germplasm resources and association analysis at adult stage. Sci Agri Sin 50 (4):625–639

Mei HX, Liu YY, Du ZW, Wu K, Cui CQ et al (2017) High-density genetic map construction and gene mapping of basal branching habit and number of flowers per leaf axil in sesame. Front Plant Sci 8:636

Melzer S, Lens F, Gennen J, Vanneste S, Rohde A et al (2008) Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana*. Nat Genet 40:1489–1492

Miao H (2014) The sesame genome project and sesame genome sequencing. In: Plant and animal genome XXII conference. 10–15th Jan., San Diego, USA

Miao H, Zhang H (2016) The Genome of *Sesamum indicum* L. In Plant and ANIMAL GENOME XXIV Conference. 9–13th, January, San Diego, USA.

Miao HM, Li C, Duan YH, Wei LB, Ju M et al (2020) Identification of a *Sidwf1* gene controlling short internode length trait in the sesame dwarf mutant *dw607*. Theor Appl Genet 133:73–86

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215–222

Sim SC, Durstewitz G, Plieske J, Wieseke R, Ganal MW et al (2012) Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. PLoS ONE 7:e40563

Spandana B, Reddy VP, Prasanna GJ, Anuradha G, Sivaramakrishnan S (2012) Development and characterization of microsatellite markers (SSR) in *Sesamum* (*Sesamum indicum,* L.) species. Appl Biochem Biotechnol 168(6):1594–1607

Surapaneni M, Yepuri V, Vemireddy L (2014) Development and characterization of microsatellite markers in Indian sesame (*Sesamum indicum* L.). Mol Breed 34 (3):1185–1200

Uncu AÖ, Gultekin V, Allmer J, Frary A, Doganlar S (2015) Genomic simple sequence repeat markers reveal patterns of genetic relatedness and diversity in sesame. Plant Genome 8(2):1–12

Uncu AO, Frary A, Karlovsky P, Doganlar S (2016) High-throughput single nucleotide polymorphism (SNP) identification and mapping in the sesame (*Sesamum indicum* L.) genome with genotyping by sequencing (GBS) analysis. Mol Breed 36(12):173

Uzun B, Çağırgan Mİ (2009) Identification of molecular markers linked to determinate growth habit in sesame. Euphytica 166:379–384

Verma P, Goyal R, Chahota RK, Sharma TR, Abdln MZ et al. (2015) Construction of a genetic linkage map and identification of QTLs for seed weight and seed size traits in Lentil (*Lens culinaris* Medik.). PLoS ONE 10:e0139666

Wang L, Zhang Y, Qi X, Gao Y, Zhang X (2012) Development and characterization of 59 polymorphic cDNA-SSR markers for the edible crop *Sesamum indicum* L. (Pedaliaceae). Amer J Bot 99:e394–e398

Wang L, Yu S, Tong C, Zhao Y, Liu Y et al (2014) Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. Genome Biol 15 (2):R39

Wang S, Chen J, Pang W, Hu Y, Chang L et al (2015) Sequence-based ultra-dense genetic and physical maps reveal structural variations of allopolyploid cotton genomes. Genome Biol 16:108–108

Wang L, Xia Q, Zhang Y, Zhu X, Zhu X et al (2016) Updated sesame genome assembly and fine mapping of plant height and seed coat color QTLs using a new high-density genetic map. BMC Genomics 17(1):1–13

Wang F, Wang C, Liu P, Lei C, Hao W et al. (2016b) Enhanced rice blast resistance by CRISPR/ Cas9-targeted mutagenesis of the ERF transcription factor gene *Os ERF922*. PLoS ONE11(4):e0154027

Wang L, Zhang Y, Zhu X, Zhu X, Li D et al (2017) Development of an SSR-based genetic map in sesame and identification of quantitative trait loci associated with charcoal rot resistance. Sci Rep 7(1):8349

Wei LB, Zhang HY, Zheng YZ, Guo WZ, Zhang TZ (2008) Developing EST-Derived microsatellites in sesame (*Sesamum indicum* L.). Acta Agron Sin 34 (12):2077–2084

Wei LB, Zhang HY, Zheng YZ, Miao HM, Zhang TZ et al. (2009) A genetic linkage map construction for sesame (*Sesamum indicum* L.). Genes Genom 31 (2):199–208

Wei WL, Qi X, Wang LH, Zhang YX, Wei H et al. (2011) Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. BMC Genomics 12(1):451

Wei WL, Zhang YX, Lv HX, Wang LH, Li DH et al. (2012) Population structure and association analysis of oil content in a diverse set of Chinese sesame (*Sesamum indicum* L.) germplasm. Sci Agri Sin 45 (10):1895–1903

Wei WL, Zhang YX, Lv H, Li DH, Wang LH et al (2013) Association analysis for quality traits in a diverse panel of Chinese sesame (*Sesamum indicum* L.) germplasm. J Integr Plant Biol 55:745–758

Wei LB, Miao HM, Li C, Duan YH, Niu JJ et al (2014) Development of SNP and InDel markers via de novo transcriptome assembly in *Sesamum indicum* L. Mol Breed 34:2205–2217

Wei X, Wang L, Zhang Y, Qi X, Wang X et al (2014) Development of simple sequence repeat (SSR) markers of sesame (*Sesamum indicum*) from a genome survey. Molecules 19(19):5150–5162

Wei X, Liu K, Zhang Y, Feng Q, Wang L et al (2015) Genetic discovery for oil production and quality in sesame. Nat Commun 6:8609

Wei W, Zhang Y, Wang L, Li D, Gao Y et al (2015) Genetic diversity, population structure, and association mapping of 10 agronomic traits in sesame. Crop Sci 56(1):331–343

Wu K, Yang M, Liu H, Ye T, Mei J et al. (2014a) Genetic analysis and molecular characterization of Chinese sesame (*Sesamum indicum* L.) cultivars using insertion-deletion (InDel) and simple sequence repeat (SSR) markers. BMC Genet 15(1):35

Wu K, Liu H, Yang M, Tao Y, Ma H et al (2014b) High-density genetic map construction and QTL analysis of grain yield-related traits in sesame (*Sesamum indicum* L.) based on rad-seqtechonology. BMC Plant Biol 14 (1):274

Wu K, Wu WX, Yang MM, Liu HY, Hao GC et al (2017) QTL mapping for oil, protein and sesamin contents in seeds of white sesame. Acta Agron Sin 43(7):1003–1011

Yang M, Yang WJ, Gao Y, Zhang YX, Zhu XD et al (2017) Quantitative trait locus mapping for sesame capsule size. Chin J Oil Crop Sci 39(6):785–793

Yepuri V, Surapaneni M, Kola V (2013) Assessment of genetic diversity in sesame (*Sesamum indicum* L.) genotypes, using EST-derived SSR markers. J Crop Sci Biotechnol 16(2):93–103

Yue W, Wei L, Zhang T, Li C, Miao H et al. (2012) Genetic diversity and population structure of germplasm resources in sesame (*Sesamum indicu*m L.) by SSR Markers. Acta Agron Sin 38(12):2286–2296

Zhang YX, Zhang XR, Hua W, Wang LH, Che Z (2010) Analysis of genetic diversity among indigenous landraces from sesame (*Sesamum indicum* L.) core collection in China as revealed by SRAP and SSR markers. Genes Genomics 32:207–215

Zhang H, Miao H, Li C, Wei L, Ma Q (2012) Analysis of Sesame Karyotype and Resemblance-near Coefficient. Bull Bot 47(6):602–614

Zhang HY, Wei LB, Miao HM, Zhang TD, Wang CY (2012) Development and validation of genic-SSR markers in sesame by RNA-seq. BMC Genomics 13:316

Zhang YX, Wang LH, Li DH, Wei WL, Gao Y et al. (2012c) Association mapping of sesame(*Sesamum indicum* L.) resistance to *Macrophomina Phaseolina* and identification of resistant accessions. Sci Agri Sin 47(3):422–430

Zhang YX, Wang LH, Xin H, Li D, Ma C et al (2013) Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. BMC Plant Biol 13(1):1–12

Zhang H, Miao H, Wei L, Li C, Zhao R et al. (2013a) Genetic analysis and QTL mapping of seed coat color in sesame (*Sesamum indicum* L.). PLoS ONE 8(5): e63898

Zhang YX, Wang LH, Li DH, Gao Y, Lv HX et al (2014) Mapping of sesame waterlogging tolerance QTL and identification of excellent waterlogging tolerant germplasm. China Agri Sci 47(3):422–430

Zhang H, Miao HM, Li C, Wei LB, Duan YH et al (2016) Ultra-dense SNP genetic map construction and identification of *SiDt* gene controlling the determinate growth habit in *Sesamum indicum* L. Sci Rep 6:31556

Zhang HY, Miao HM, Ju M (2019) Potential for adaptation to climate change through genomic breeding in sesame. In: Kole C (ed) Genomic designing of climate-smart oilseed crops. Springer, Cham, Switzerland, pp 374–376

Zhao J, Artemyeva A, Del Carpio DP, Basnet RK, Zhang N et al (2010) Design of a *Brassica rapa* core collection for association mapping studies. Genome 53:884–898

Zhao X, Han Y, Li Y, Liu D, Sun M et al (2015) Loci and candidate gene identification for resistance to *Sclerotinia sclerotiorum* in soybean (*Glycine max* L. Merr.) via association and linkage maps. Plant J 82:245–255

Zhou R, Dossa K, Li D, Yu J, You J et al (2018) Genome-wide association studies of 39 seed yield-related traits in sesame (Sesamum indicum L.). Int J Mol Sci 19 (9):2794

Cytological Details of Sesame

9

Hongmei Miao, Ruihong Zhao, Qin Ma, and Haiyang Zhang

Abstract

The chromosome morphology, chromosome structure, karyotype, and the chromosome set relationship in *Sesamum* are presented in this chapter. Besides, the modified chromosome preparation and application of BAC-FISH techniques are briefly introduced at the same time. The 13 chromosomes of *S. indicum* have been identified and marked using 27 BACs. A high-density cytogenetic map for sesame is constructed with 210 BACs using successive double-colored BAC-FISH. A tetraploid species *S. radiatum* was initially determined using the cytogenetic data. The main progress in cytogenetical research in *Sesamum* species provides the theoretical basis for further genomic research in *Sesmaum*.

9.1 Introduction

Sesame is an ancient crop and occupied an important position in the phylogenic system among angiosperms. There are about 36 species in *Sesamum* according to the index Kewensis (Joshi 1961). The chromosome research in sesame dates back to 1920s. Morinaga et al. (1929) reported that the chromosome number of the cultivated sesame (*S. indicum*) is $2n = 26$ for the first time. Kumar and Abraham (1941) found the secondary association in sesame. Meanwhile, some scientists inferred that the basic chromosome number may be 8, based on the secondary association and meiosis in sesame (Joshi 1961), although this conclusion has not been proved.

In *Sesamum*, cytogenetical characters, such as the chromosome number, satellite chromosome number, chromosome structure, and karyotype of the 36 species, are diverse, as well as the botanical and morphological characteristics. To date, the *Sesamum* species are divided into three categories according to the somatic chromosome number, i.e., $2n = 26$ (e.g., *S. indicum*, *S. alatum*), $2n = 32$ (e.g., *S. angolense*, *S. latifolium*), and $2n = 64$ (e.g., *S. radiatum*, *S. schinizianum*), even though the chromosome characters in some species are still unknown (Nimmakayala et al. 2011; Bedigian 2015). By the 1980s, Giemsa staining and other karyotype analysis techniques were established and applied in sesame (Zhan et al. 1987, 1988, 1990; He et al. 1994; Zhang et al. 2012). These techniques improved the analyses of chromosome structure and karyotype in *Sesamum*. Of note, the establishment of the fluorescence in situ hybridization (FISH) and bacterial artificial chromosome (BAC) FISH techniques in recent years has really improved the cytological studies of sesame (Liu et al. 2013,

H. Miao · R. Zhao · Q. Ma · H. Zhang (✉)
Henan Academy of Agricultural Sciences, Henan Sesame Research Center, Zhengzhou, People's Republic of China
e-mail: zhanghaiyang@zzu.edu.cn

H. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-98098-0_9

2014; Zhao et al. 2018a). Based on the successive BAC-FISH hybridization system, the 13 chromosome pairs of *S. indicum* were differentiated and numbered with specific BAC probes for the first time (Zhao et al. 2018a). Construction of the high-density cytogenetic map of sesame provides the chromosome structure information for genome assembly and genomics research in *Sesamum*.

9.2 Chromosome Characters of *S. Indicum* and Wild *Sesamum* Species

9.2.1 Chromosome Morphology and Karyotype of *S. Indicum*

The chromosomes of *S. indicum* ($2n = 26$) are of small size. In the 1980s, the somatic chromosomes of the cultivated sesame were observed using the traditional chromosome press technique. In sesame var. Beijing Bawangbian, the chromosome absolute length varied from 1.6 to 3.4 μm, while the ranges of the relative length and arm ratios were 5.16–10.97%, and 1.3–3.0, respectively (Zhan et al. 1987) (Fig. 9.1a). The karyotype formula reported was $2n = 2x = 5$ m + 8sm (1 SAT). Here 'm' indicates metacentric type ($1.0 \leq AR < 1.7$), and 'sm' indicates sub-metacentric type ($1.7 \leq AR < 3.0$) with various values of the relative length ratios of the long arm to the short arm of the chromosome (AR), according to preexisting criteria (Stebbins 1971; Chen and Li 1985; Paszko 2006). 'SAT' indicates the satellite chromosome. In the chromosome set of var. Beijing Bawangbian, one pair of satellite chromosome was found (Fig. 9.1b).

Considering that the somatic chromosomes cannot be differentiated just based on the morphological characters, Zhan et al. (1988) applied the chromosome Giemsa banding technology, especially the barium hydroxide, salt, and Giemsa (BSG) method in sesame chromosome research. The results reflected that the C-band form of the cultivated sesame var. Beijing Bawangbian was $2n = 26$ = CITOW type = 8 $C + 8CI_+ + 4CT_+ + 2CI_+ + 2O + 2$ W in which, 'C' indicates the belt located in centromere region; 'I' indicates the belt located between centromere and chromosome terminal; 'T' indicates the belt in the terminal region; 'N' indicates the belt in nucleolar constriction region; 'O' means no belt; and 'W' indicates the belt on the whole chromosome. The belts in long arms of 'I' and 'T' type chromosomes are signed with '+'. For the variety Beijing Bawangbian, the $2CI_+$ in the chromosomes possessed the satellite structure but no N belt. The authors believed that the homologous chromosomes in sesame could be identified, based on the relative length, centromere position, and C-banding types (i.e., C, I, T, O, and W) (Zhan et al. 1988).

Subsequently, Zhan et al. (1990) analyzed eight typical sesame cultivars with white or other seed colors. The chromosome form was $2n = 2x = 16$ m (1SAT) + 10sm. The chromosome Giemsa C-belt type of white sesame was $2n = 26$ = CITOW type = $8C + 10CI_+ + 4CT_+ + 2O + 2$ W. The formula of other varieties with the various seed colors was $2n = 26$ = CITOW type = $8C + 6CT_+ + 2O + 2W$. Meanwhile, He et al. (1994) compared ten sesame varieties and reported the same number and similar karyotype to the previous reports. However, four satellite chromosomes of the 26 chromosomes in sesame were found. The formula of sesame was $2n = 26 = 12m + 12sm$ (4SAT) + 2st. Of 26 chromosomes, two presented sub-telocentric type (st). From the above results, we can find that the exact number of satellite chromosomes in sesame has not been determined using conventional cytogenetical technology.

Many reports prove that the morphology of chromosomes in sesame is affected significantly by chromosomes specimen preparation technology (He et al. 1994; Zhang et al. 2012; Liu et al. 2014; Zhao et al. 2018a, b). The satellite region of the sesame chromosomes sometimes presents indistinct or invisible mainly owing to unavoidable irregularities of pretreatment (Liu et al. 2014). Therefore, to decrease the effect of chromosome specimen preparation, the Chinese scientists established the modified chromosome

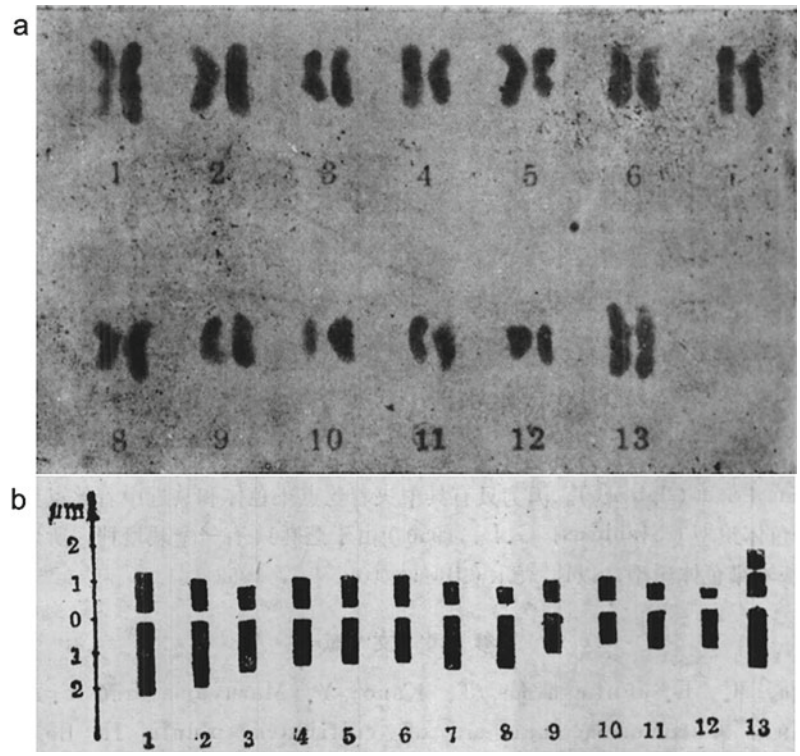


Fig. 9.1 Chromosome morphology and karyotype of the cultivated sesame. **a** Chromosome morphology of sesame. **b** Karyotype of sesame (Cited from Zhan et al. (1987))

preparation and chromosome press techniques (Zhang et al. 2014; Zhao et al. 2018a, b). Differing from the conventional pressing technique, the press technique involves the modified chromosome preparation, the wall degradation hypotonic method, and the specimen press method. The steps of applying cover plates and removing cover plates are deleted. Using the new preparation technique, the chromosomes spread more naturally and tightly on the high-quality slides and are suitable for morphological analysis and further cytogenetical research (Tran et al. 2017; Zhao et al. 2018a, b). The disturbance on the BAC hybridization signal from the cytoplasm can significantly reduce.

From thousands of chromosome specimens of sesame var. Yuzhi 11 using Giemsa staining, a chromosome specimen with three pairs of satellite chromosomes has been observed under microscope (Fig. 9.2a). By the way, five satellite chromosomes (indicated by arrows) are also found in some other chromosome specimens (Fig. 9.2b) (Unpublished data, Haiyang Zhang). To date, FISH hybridization using 45S rDNA proved that there are three pairs of satellite chromosomes in sesame (Zhao et al. 2018a, b).

In order to systematically reflect the chromosome karyotype characters of *S. indicum*, Zhang et al. (2012) compared the chromosome types and karyotypes of eight diploid and four artificial tetraploid materials of *S. indicum*, and two wild species (*S. radiatum* Schum and Thonn and *S. schinzianum* Asch) using the traditional chromosome press technique (Fig. 9.3). The results indicated that S. *indicum* centromere types include m, sm, and st; the average arm ratio (AAR) ranged from 2.11 to 2.25, asymmetric karyotype coefficients (AKC) ranged from 66.05 to 66.99%, and the karyotypes belonged to 2B, 3A, and 3B, respectively (Fig. 9.3a, b). In contrast, *S. radiatum* and *S. schinzianum* exhibited only M and m centromere types; the AAR was 1.09 and 1.08, respectively; AKC was 52.12 and 51.68%, respectively. The karyotypes of *S. radiatum* and *S. schinzianum* were 1B and 1A, respectively (Fig. 9.3c, d). The resemblance near

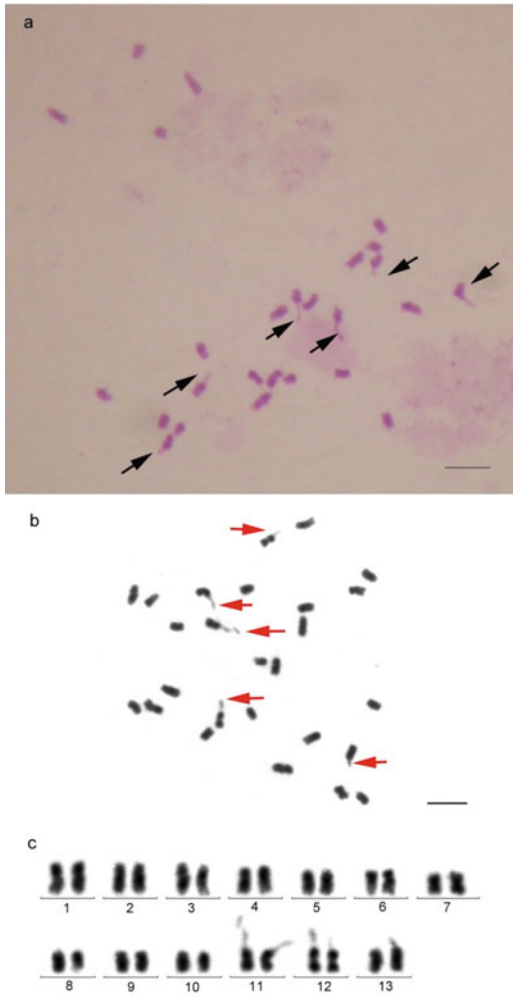


Fig. 9.2 Chromosome morphology and karyotype of the cultivated sesame Yuzhi 11. **a** Three pairs of satellite chromosomes (indicated by arrows) are found in a chromosome specimen in var. Yuzhi 11 using Giemsa staining. **b** Five satellite chromosomes (indicated by arrows) are found in another chromosome specimen of var. Yuzhi 11. **c** Karyotype of Yuzhi 11 chromosomes with three pairs of satellite chromosomes (Photographs provided by Ruihong Zhao)

coefficient of *S. indicum* and the wild species was 0.027–0.107, indicating the remote genetic relationship.

The karyotype analyses in many reports confirm that most chromosomes in *S. indicum* are m and sm, with average lengths of 1.177–2.128 μm. Zhao et al. (2018a) analyzed the karyotype of sesame (var. Yuzhi 11) using the Metasystems Isis software (MetaSystems, Altlussheim, Germany) using the new chromosome preparation technique (Table 9.1). In this chromosome set of var. Yuzhi 11, SiChr.11, SiChr.12, and SiChr.13 are satellite chromosomes. For sesame, the chromosomes belong to the 3A type. The length of 13 chromosomes ranges from 1.32 to 2.13 μm. The asymmetry index is 63.1%. The karyotype formula is $2n = 2x = 26 = 12m + 14sm$, consisting of 12 M1 ($0.76 \leq$ I.R.L. ≤ 1.00) (I.R.L. = ratio of a chromosome relative length to the average relative length of 13 chromosomes), 12 M2 ($1.01 \leq$ I.R.L. ≤ 1.25), and 2S (I.R.L. < 0.76) type chromosomes (Table 9.1). The relative length of 13 chromosomes ranges from 5.73–9.22%. The length ratio of the longest to the shortest chromosome is 1.61.

9.2.2 Chromosome Morphology and Karyotype of Wild *Sesamum* Species

In *Sesamum* species, the somatic chromosome number of most species has been determined (Nimmakayala et al. 2011). Benson et al. (2014) reported the somatic chromosome number of four wild species, namely *S. calycimum* ($2n = 32$), *S. angolense* ($2n = 32$), *Sesamum latifolium* ($2n = 32$), and *Sesamum* spp. ($2n = 32$) in Kenya. To reflect the chromosome evolution of *Sesamum*, we perform the chromosome analysis using the six wild species covering the three chromosome categories of 26, 32, and 64. Yuzhi 11 is used as the representative of the cultivated sesame, as the genome of Yuzhi 11 has been sequenced (Unpublished data, Haiyang Zhang).

As shown in Fig. 9.4, *S. alatum* ($2n = 26$, Fig. 9.4b) has the same chromosome number with *S. indicum* (var. Yuzhi 11, Fig. 9.4a). *S. latifolium*, *S. angolense*, *S. calycinum*, and *S. angustifolium* possess 32 chromosomes (Fig. 9.4c–f), while *S. radiatum* has 64 (Fig. 9.4g). Interestingly, the number of satellite chromosomes (indicated by red arrows) of the seven species varies.

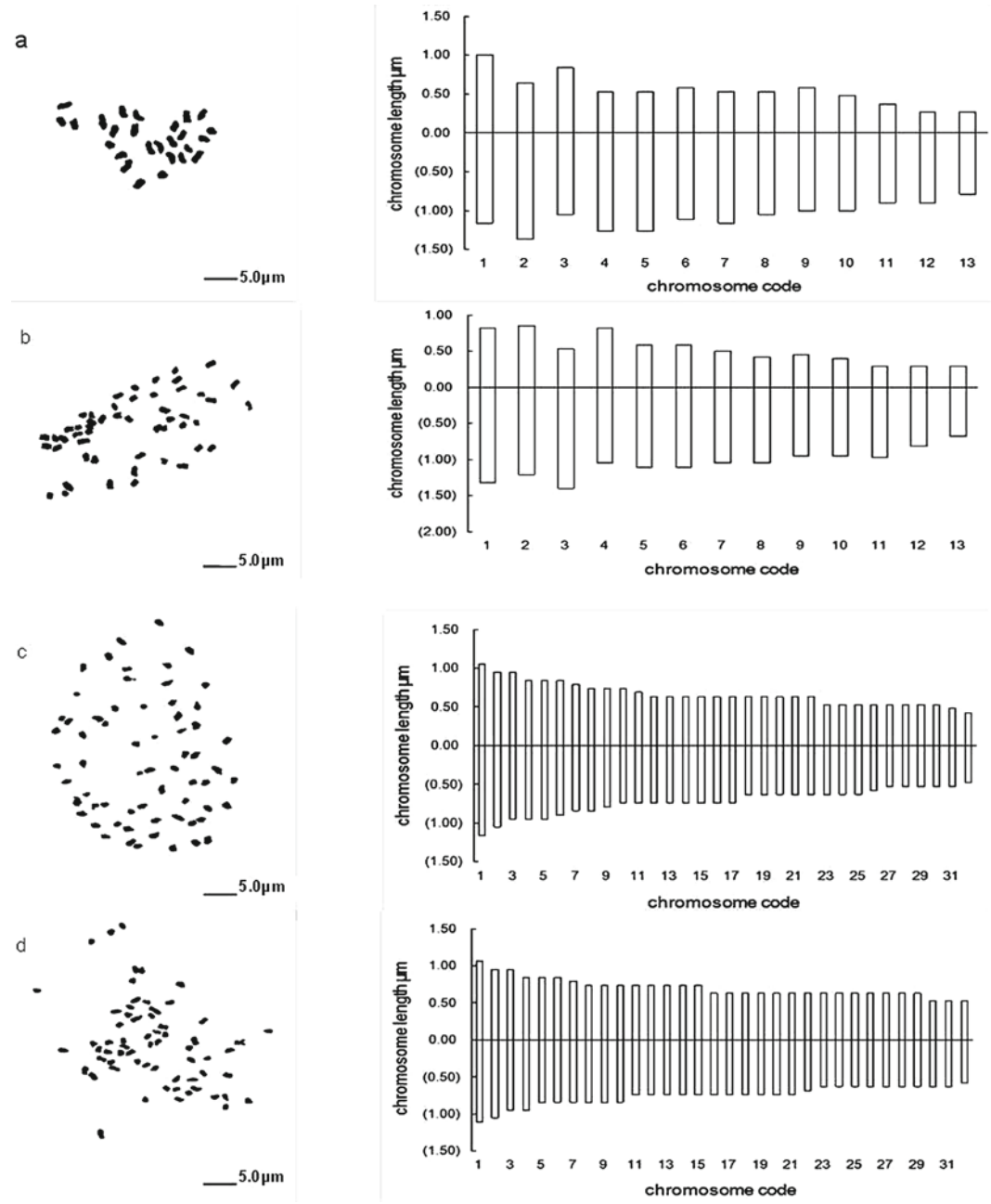


Fig. 9.3 Chromosome types and karyotype idiogramms of the cultivated sesame and two wild species of *Sesamum*. **a** Yuzhi 11 ($2n = 2x = 26$); **b** TetraploidYuzhi 4 ($2n = 4x = 52$). **c** *S. radiatum* ($2n = ?x = 64$). **d** *S. schinzianum* ($2n = ?x = 64$). (Modified from Zhang et al. (2012))

In the seven species, the maximum and the minimum total somatic chromosome length are 85.507 μm in *S. radiatum*, and 30.597 μm in *S. indicum* (var. Yuzhi 11), respectively (Table 9.2) (Unpublished data, Haiyang Zhang). Of the six wild species, *S. radiatum* has the lowest average length of chromosomes (1.336 μm), while *S. calycinum* has the highest average value

Table 9.1 Chromosome karyotype analysis of sesame variety Yuzhi 11[a]

Chromosome no	Absolute length (μm)	Relative length	Relative length of short arm	Relative length of long arm	I.R. L	Centromere type	Chromosome length type
SiChr.1	2.13	9.22	3.13	6.09	1.20	Sm	M2
SiChr.2	2.13	9.22	3.26	5.97	1.20	Sm	M2
SiChr.3	2.10	9.07	4.24	4.84	1.18	M	M2
SiChr.4	2.08	9.00	4.01	4.99	1.17	M	M2
SiChr.5	1.87	8.11	2.79	5.32	1.05	Sm	M2
SiChr.6	1.87	8.10	3.09	5.00	1.05	M	M2
SiChr.7	1.71	7.38	3.13	4.25	0.96	M	M1
SiChr.8	1.53	6.61	2.56	4.05	0.86	M	M1
SiChr.9	1.53	6.63	2.65	3.97	0.86	M	M1
SiChr.10	1.32	5.73	1.45	4.28	0.75	st	S
SiChr.11	1.66	7.18	2.38	4.80	0.93	sm	M1
SiChr.12	1.61	6.98	2.07	4.91	0.91	sm	M1
SiChr.13	1.57	6.78	2.10	4.68	0.88	sm	M1

[a]Modified from Zhao et al. (2018a)

(1.664 μm). Five chromosome parameters indicate that the wild species are different with each other. In *S. radiatum*, the value of LC-SC (variation between the largest and shortest chromosomes) ranges from 2.410 to 3.046, while *S. alatum* has high LC-SC value of 4.747. Differing from *S. indicum*, four wild species including *S. latifolium*, *S. angolense*, *S. calycinum*, and *S. angustifolium* only contain types of L, M2, and M1.

In *sesamum*, there exist three centromere types, i.e., metacentric (m) ($1.0 \leq AR < 1.7$), sub-metacentric (sm) ($1.7 \leq AR < 3.0$), and sub-telocentric (st) ($3.0 \leq AR < 7.0$). However, *S. radiatum* lacks sm and st centromeres, but possesses the specific M centromere (AR = 1.0), indicating that the chromosomes are relatively even. Karyotype AI (asymmetry index) is an indicator to reflect the symmetric character of a chromosome set. Of the seven species, *S. radiatum* has the lowest AI (51.43%), while *S. angolense* has the highest (69.30%). Compared with the other six species, *S. radiatum* chromosomes also exhibit the lowest SDcl (standard deviation of chromosome relative length) (0.49) and CVcl (variation coefficient of chromosome length) (0.035) and highest ACI (average centromeric index) (0.48). The results suggest that the *S. radiatum* has a conserved and distinct karyotype (Unpublished data, Haiyang Zhang).

9.2.3 Distribution of 45S and 5S rDNAs in *Sesamum*

Satellite chromosomes refer to the chromosomes containing 45S (18S-5.8S-25S) ribosomal DNA (rDNA) repeat sequences. The production of satellite chromosomes would result from the fusion of two chromosomes (Wang et al. 2016). Thus, the characteristics of satellite chromosomes can reflect the evolution of a species. However, there are no reports about the characters analyses of satellite chromosomes in *Sesamum* till now.

To reflect the distribution of satellite chromosomes in *Sesamum*, we performed FISH hybridization using 45S (in green) and 5S (in red) rDNA probes in the seven tested *Sesamum* species (Fig. 9.5) (Unpublished data, Haiyang Zhang). The red color signal released from the Tetramethyl-Rhodamine-5-dUTP, while the green color signal is released from fluoresce. Probes of rDNA repeats (45S rDNA and 5S rDNA) were

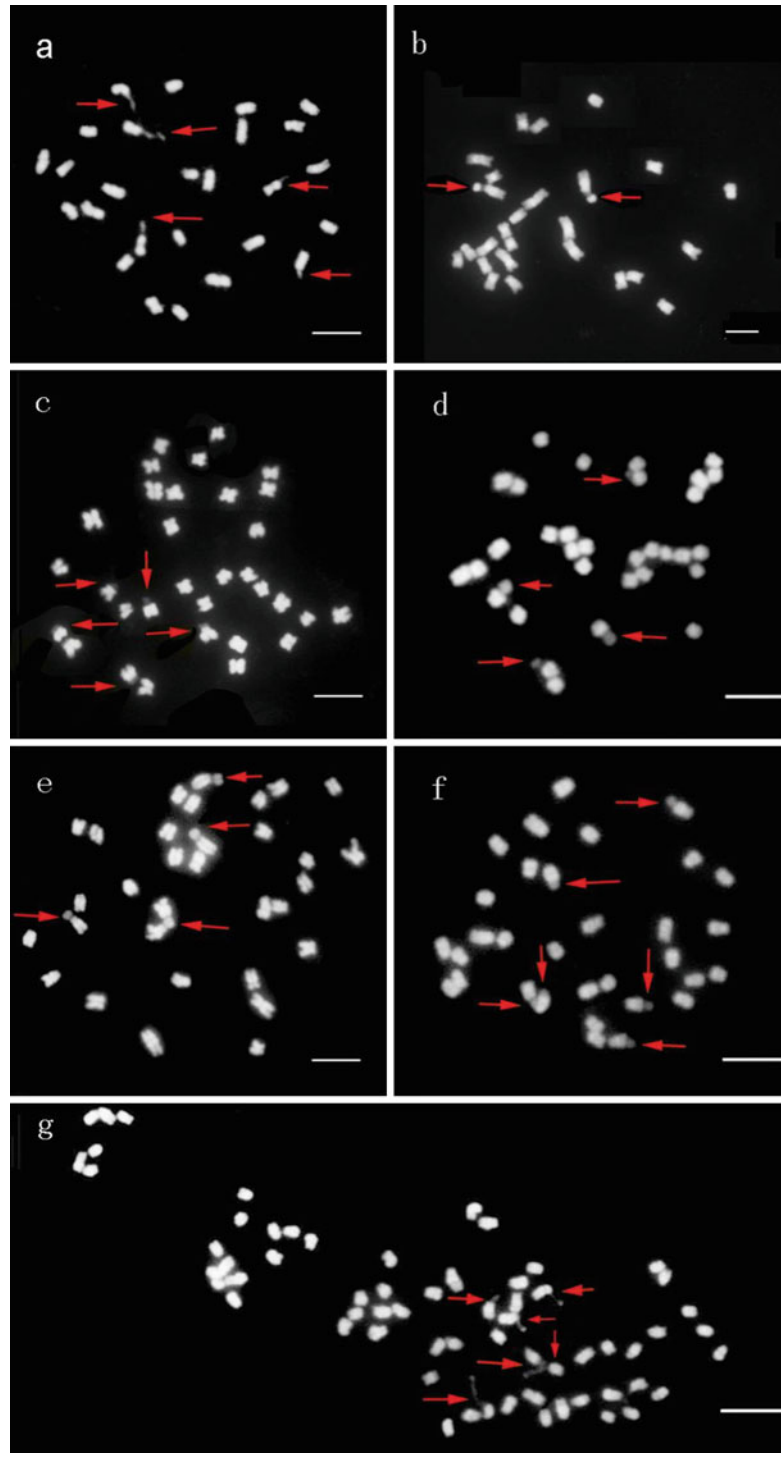


Fig. 9.4 Morphological comparison of the somatic chromosomes of the cultivated sesame and the six *Sesamum* species. **a** *S. indicum* (*var.* Yuzhi 11); **b** *S. alatum*; **c** *S. latifolium*; **d** *S. angolense*; **e** *S. calycinum*; **f** *S. angustifolium*; **g** *S. radiatum*. Bar = 5 μm. Red arrow indicates the satellite chromosome (Photographs provided by Ruihong Zhao)

Table 9.2 Information of basic chromosome parameters of the seven *Sesamum* species

Species	Material name	Chromosome number	Constitution of chromosome set	Karyotype type	Chromosome relative size					Asymmetry index	
					AI (%)	LC-SC	SDcl	LC/SC	CVcl	ACI	CVci
S. indicum L	Yuzhi 11	26	12M2 + 12M1 + 2S	3A	65.67	3.430	1.017	1.565	0.132	0.336	0.265
S. alatum Thonn	3651	26	12M2 + 12M1 + 2S	3A	65.67	3.430	1.017	1.565	0.132	0.336	0.265
S. latifolium G	Ken1	32	2L + 10M2 + 20M1	3A	69.30	2.982	0.847	1.570	0.136	0.312	0.308
S. angolense W	K16	32	4L + 10M2 + 18M1	3A	67.34	3.046	0.848	1.576	0.136	0.330	0.286
S. calycinum W	Ken8	32	4L + 6M2 + 22M1	2A	66.59	3.492	0.992	1.698	0.159	0.343	0.278
S. angustifolium Oliver Engl	G01	32	2L + 12M2 + 18M1	2A	66.07	2.966	0.819	1.420	0.131	0.345	0.243
S. radiatum Schumach. & Thonn	G02	64	4L + 28M2 + 28M1 + 4S	1B	51.43	2.410	0.497	2.15	0.159	0.485	0.035

AI Asymmetry index; *LC-SC* Variation between the largest and shortest chromosomes; *SDcl* Standard deviation of chromosome relative length; *LC/SC* Length ratio between the largest chromosome and the shortest chromosome; *CVcl* Variation coefficient of chromosome length; *ACI* Average centromeric index; *CVci* Variation coefficient of centromeric index (Provided by Haiyang Zhang)

synthesized according to the conservative sequences of *Arabidopsis thaliana* reported by Chen et al. (2013a, b). The idiograms of the 45S and 5S rDNAs distribution pattern of 45S and 5S in chromosome sets are drawn (Fig. 9.6).

Similar to *S. indicum* (Fig. 9.5a), *S. latifolium* (Fig. 9.5c), and *S. angustifolium,* (Fig. 9.5f) all have three pairs of signals hybridized by 45S rDNA probe. *S. radiatum* has the highest satellite signal pairs (five) hybridized by 45S rDNA (Fig. 9.5g), while *S. alatum*, *S. angolense* (Fig. 9.5d), and *S. calycinum* (Fig. 9.5e) present the lowest 45S rDNA pairs (two) (Fig. 9.5b).

For 5S rDNA hybridization, the highest signal pair also presents in *S. radiatum* (five), followed by *S. calycinum* (three) and *S. angustifolium* (three). Interestingly, two 5S rDNA loci are specifically observed on both arms of an *S. alatum* chromosome (Fig. 9.5b).

In comparison with the hybridization signals of both probes, we detect that *S. calycinum* (Fig. 9.5e) has more 5S rDNA loci (three) than 45S rDNA (two). In *S. alatum*, *S. calycinum*, and *S. radiatum*, partial 45S and 5S rDNA are coupled and co-localized in specific chromosome pairs. Meanwhile, *S. angolense* (Fig. 9.5d) and

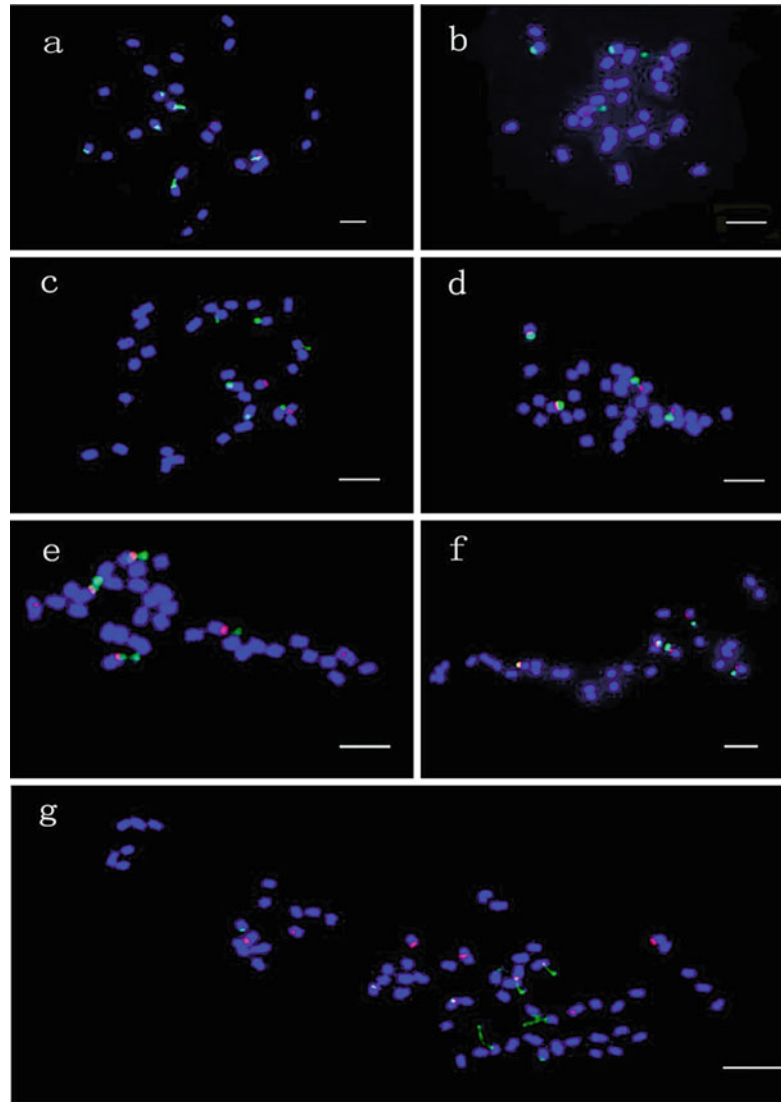


Fig. 9.5 Florescence in situ hybridization of metaphase chromosomes of the seven *Sesamum* species using 45S and 5S rDNA probes. **a** *S. indicum* (*var.* Yuzhi 11); **b** *S. alatum*; **c** *S. latifolium*; **d** *S. angolense*; **e** *S. calycinum*; **f** *S. angustifolium*; **g** *S. radiatum*. Bar = 5 μm. The probes are labeled with tetramethyl-rhodamine-5-dUTP (red color signal) or fluorescein-12-dUTP (green color signal). Green signal is hybridized by 45S rDNA probe. Red signal is hybridized by 5S rDNA probe (Photographs provided by Hongmei Miao)

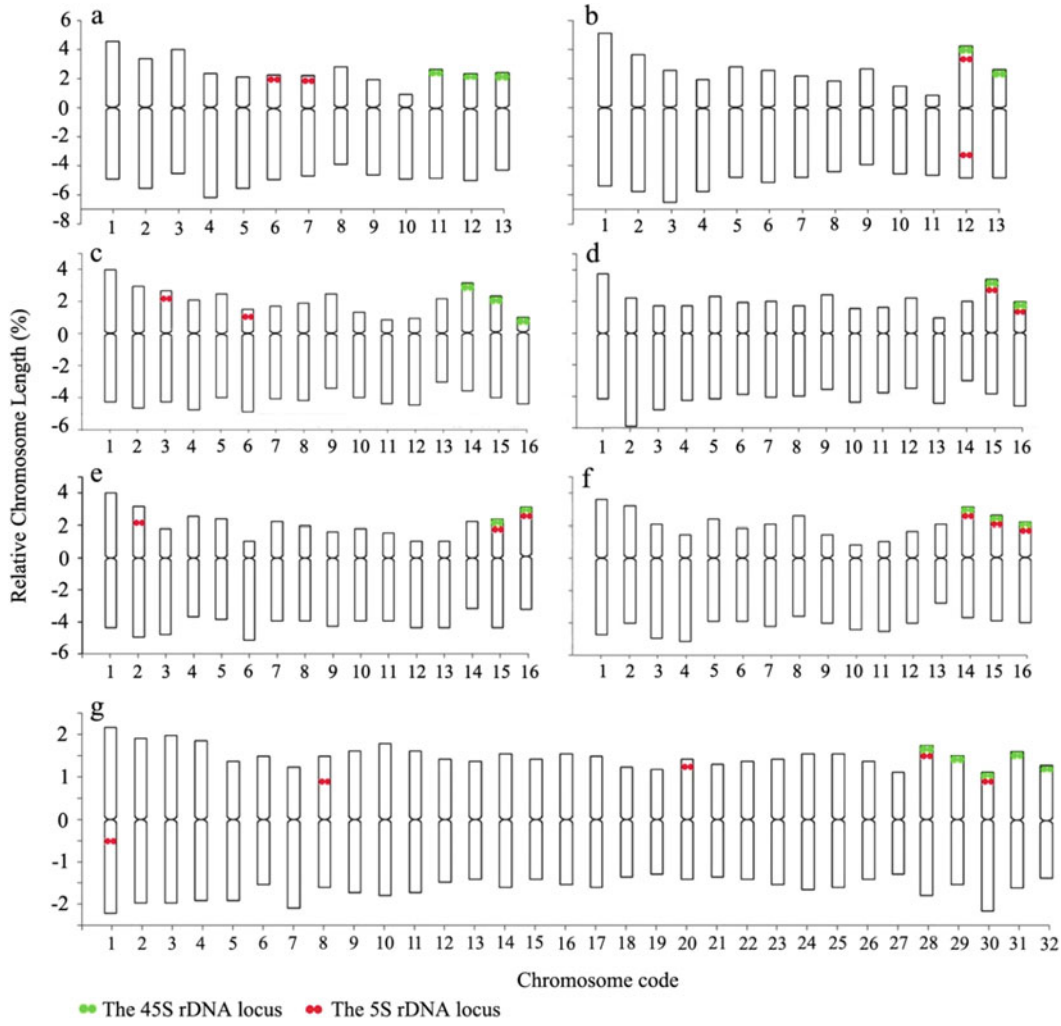


Fig. 9.6 Idiograms of the distribution pattern of 45S and 5S rDNAs in chromosome sets of the seven *Sesamum* species. **a** *S. indicum* (var. Yuzhi 11); **b** *S. alatum*; **c** *S. latifolium*; **d** *S. angolense*; **e** *S. calycinum*; **f** *S. angustifolium*; **g** *S. radiatum* (Provided by Hongmei Miao)

S. angustifolium (Fig. 9.5f) exhibit stringent co-localization of all 45S and 5S rDNAs.

In addition, of the seven *Sesamum* species, most 45S and 5S rDNA loci are mapped in short chromosome arms, except for *S. radiatum* and *S. alatum*. Specifically, 45S rDNA fluorescence signals are always detected at terminal short arms, while 5S rDNA signals are terminally or subterminally located on the short arm. The results reflect that the rDNA distribution in chromosome sets of *Sesamum* varies. Therefore, the diverse cytogenetical characters of the *Sesamum* species can be used as the references to differentiate the unknown species in *Sesamum*.

9.3 Chromosome Identification of *S. Indicum*

9.3.1 Construction of BAC-FISH System

The above studies of chromosome morphology indicate that sesame chromosomes are too small,

and most are similar with each other. Moreover, the morphology and the size of chromosome set are instable and susceptible to environmental factors during specimen preparation. In order to accurately differentiate the 13 chromosome pairs in *S. indicum*, we initially established the BAC-FISH system for sesame (Zhang et al. 2014; Zhao et al. 2018a).

In the BAC-FISH system, the BAC library (CopyConyrol™ pCC1BAC™) of *S. indicum* (var. Yuzhi 11) is applied (Zhang et al. 2013; Miao 2014; Zhao et al. 2018a). The BAC library contains 57,600 BAC clones with the average insert size of 85 kb. Thousands of BAC clones are chosen randomly from the library and applied for hybridization (Fig. 9.7).

As shown in Fig. 9.7, one pair, two pairs, three pairs, and even 13 pairs of BAC hybridization signals are successfully hybridized in sesame chromosomes using the single specific BAC probes with the high-efficient BAC-FISH system. The BACs generating more than one pair of signals may indicate that there are different degrees of repetition DNA in their inserts. Further statistics results indicate that 36 (9%) of the 400 BAC clones generate more than one pair of stable hybridization signals, and 64 (16%) generate a single stable signal pair. Compared with the hybridization ratios in *Genlisea margaretae* (11%) (Tran et al. 2017) and sugar beet (9.84%) (Paesold et al. 2012), the hybridization ratio in sesame is higher. This hybridization technology provides a useful method for chromosome identification and structural analysis in sesame.

In addition, we establish the successive (double color) BAC hybridization technique for sesame in order to increase hybridization efficiency. For example, we perform the eight rounds of hybridizations with 16 BACs locating on the same chromosome set (Fig. 9.8). Through eight successive hybridizations, the morphology of the chromosomes is still clear, and the hybridization signal is stable.

Subsequently, we integrated the eight images together (Fig. 9.9). The result shows that the 16 BACs are distributed into ten pairs of

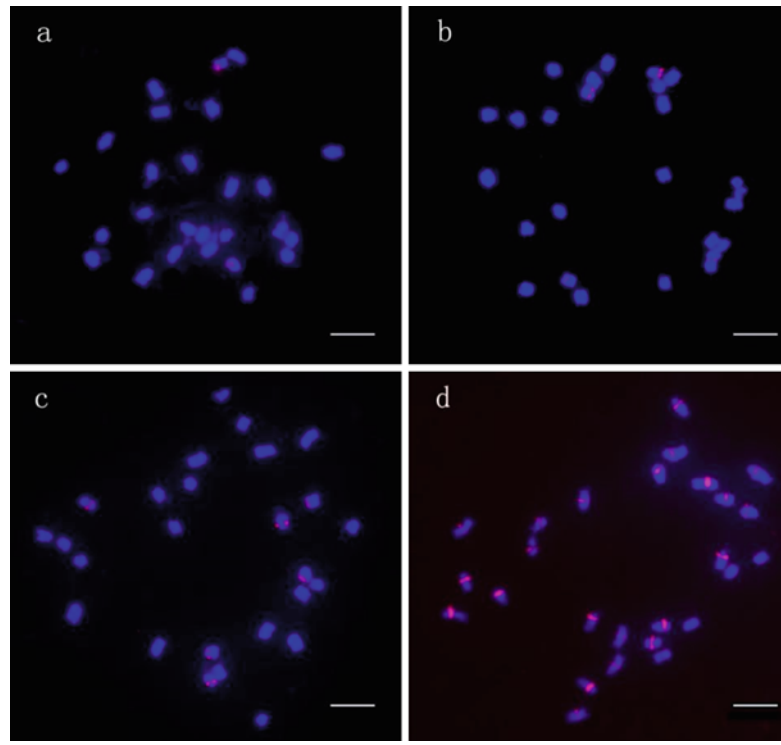


Fig. 9.7 Fluorescence in situ hybridization of different BAC probes on sesame chromosomes. **a** Hybridization of probe 3R with one signal pair. **b** Hybridization of probe 144R with two signal pairs. **c** Hybridization of probe 37R with three signal pairs. **d** Hybridization of probe B409R with 13 signal pairs. Bar = 5 μm. Red arrow indicates the fluorescence signal generated using tetramethyl-rhodamine-5-dUTP. 'R' in the probe names refers to the red fluorescent label (Modified from Zhao et al. (2018a))

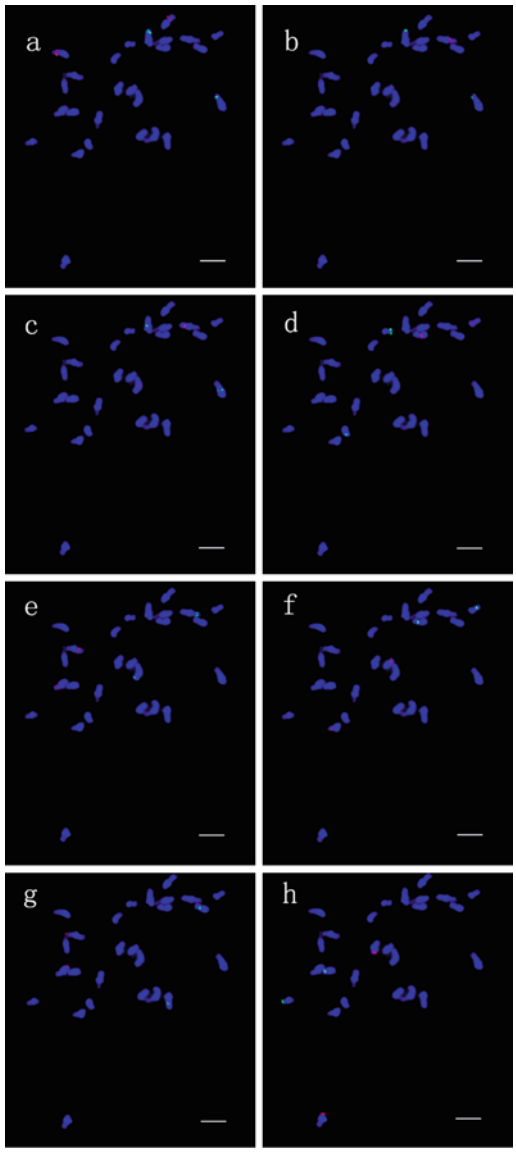


Fig. 9.8 Eight rounds of successive hybridization in the same chromosome set. **a** Hybridization of probe 48G and 40R; **b** Hybridization of probe 61G and 79R; **c** Hybridization of probe 65G and 58R; **d** Hybridization of probe 101G and 3R; **e** Hybridization of probe 121G and 32R; **f** Hybridization of probe 125G and 137R; **g** Hybridization of probe 138G and 24R; **h** Hybridization of probe 39G and 28R; Bar = 5 μm. In probe names, 'G' refers to the green fluorescent label, and 'R' refers to the red fluorescent label (Provided by Hongmei Miao)

chromosomes. The specific positions and relationships of the 16 BACs in chromosome set could be identified.

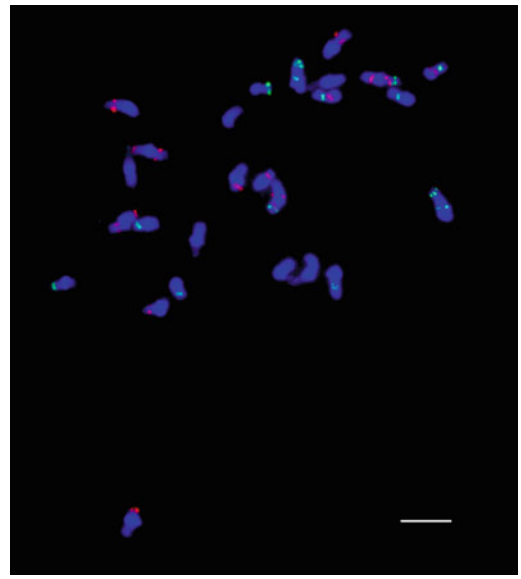


Fig. 9.9 Integration images from eight-round hybridizations on sesame chromosomes. Bar = 5 μm (Provided by Hongmei Miao)

To date, the maximum round of successive BAC-FISH hybridization is up to 12 in sesame (data not shown). The successive hybridization technology made it easier to group a great deal of BAC probes into the 13 pairs of sesame chromosome set. Thus, identifying the individual chromosome in sesame using the specific BAC markers is possible.

9.3.2 Identification of Sesame Chromosomes Using Special BAC Markers

Differentiating each chromosome is requisite to impel the chromosomal and genomic analysis for sesame. Considering high somatic chromosomes in sesame, we determined the two-step BAC-FISH strategy using the successive BAC-FISH techniques to identify the 13 chromosome pairs in sesame (Zhao et al. 2018a). As described in the above sections, we screened hundreds of the specific BACs with one pair of hybridization signals from the BAC library using BAC-FISH technique. In the first step, the 13 chromosome pairs were divided into two groups using 45S

rDNA probe (Fig. 9.10a).Then we successfully performed a successive eight-round double-color FISH using 14 BACs with a single hybridization signal pair (Fig. 9.10a–h).

As a result, the ten chromosome pairs in group I are differentiated (Fig. 9.11), while group II includes three chromosome pairs only with satellite DNA.

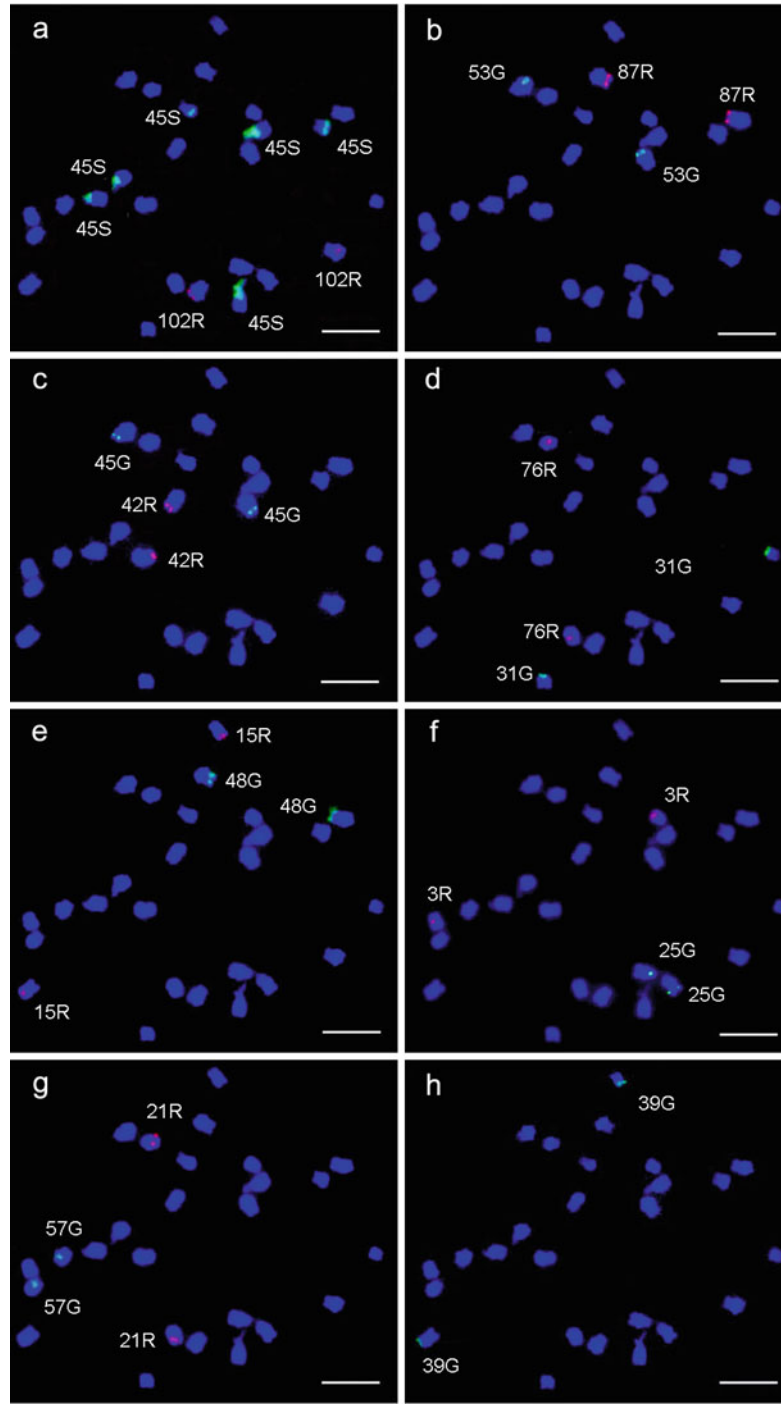


Fig. 9.10 Eight-round successive fluorescence in situ hybridization in sesame. **a** Hybridization of 102R and 45S rDNA probes (green fluorescent labeling). **b** Hybridization of 87R and 53G probes. **c** Hybridization of 42R and 45G probes. **d** Hybridization of 76R and 31G probes. **e** Hybridization of 15R and 48G probes. **f** Hybridization of 3R and 25G probes. **g** Hybridization of 21R and 57G probes. **h** Hybridization of probe 39G. In probe names, 'G' and 'R' in probe names refer to the green and red fluorescent labeling, respectively. Bar = 5 μm (Modified from Zhao et al. (2018a))

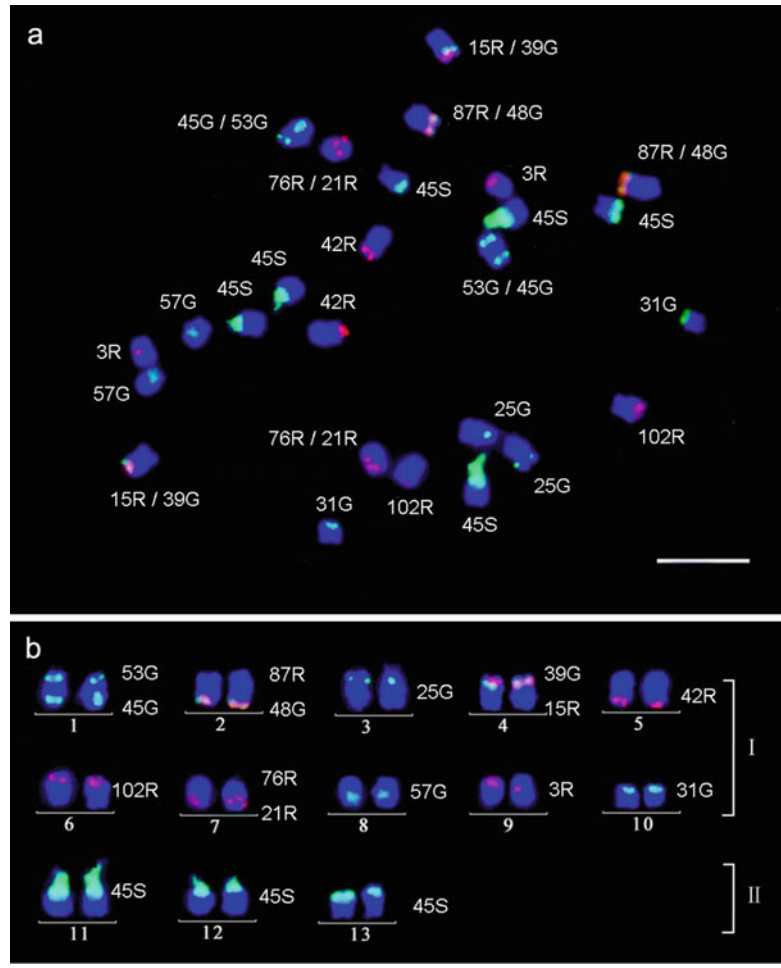


Fig. 9.11 Identification of the ten normal chromosomes (without satellite DNAs) in sesame. **a** Integration of images from eight-round hybridizations. **b** Karyotype analysis of chromosomes hybridized with the 14 BAC probes and 45S rDNA probe. 'G' and 'R' in probe names refer to the green and red fluorescent labeling, respectively. Bar = 5 μm (Modified from Zhao et al. (2018a))

In the second step, in order to differentiate group II chromosome pairs, we performed a six-round successive hybridization using 11 BAC probes and 45S rDNA (Fig. 9.12). In the chromosome set, the three chromosomes 11', 12', and 13' are marked by three BAC probes and the 45S rDNA probe (Fig. 9.13).

Subsequently, we integrated the BACs hybridized in chromosome group I in Fig. 9.11 and group II in Fig. 9.13. As shown in Fig. 9.14, the 13 chromosome pairs of *S. indicum* are specifically marked using the 17 specific BACs. In particular, Chr. 11, Chr. 12, and Chr. 13 chromosome pairs locate with satellite DNAs. The somatic chromosomes in sesame are successfully identified for the first time. The results supply the foundation for cytogenetic map construction and genome assembly in sesame.

9.3.2.1 Construction of a High-Density Cytogenetic Map in Sesame

A cytological map reflects the exact physical position of the chromosome fragments in chromosome set. In order to explore the positions of sesame DNA fragments in chromosomes and to supply the molecular markers for genome assembly, constructing a high-density cytogenetic map is necessary for sesame. Before establishing the cytogenetic map for sesame, a total of 210 BACs with the single and stable hybridization signal pair have been located in the 13 chromosomes using

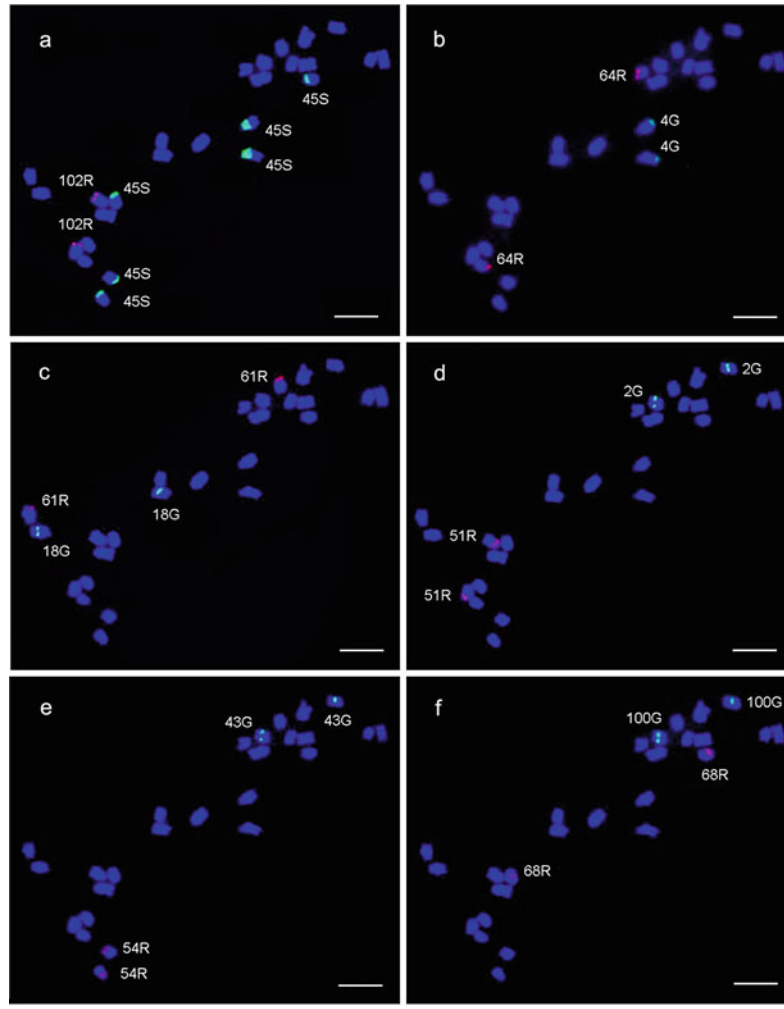


Fig. 9.12 Six-round successive fluorescence in situ hybridization in sesame. **a** Hybridization of 102R and 45S rDNA probes (G). **b** Hybridization of 64R and 4G probes. **c** Hybridization of 61R and 18G probes. **d** Hybridization of 51R and 2G probes. **e** Hybridization of 54R and 43G probes. **f** Hybridization of 68R and 100G probes. 'G' and 'R' in probe names refer to green and red fluorescent labeling, respectively. Bar = 5 μm (Modified from Zhao et al. (2018a))

the successive BAC-FISH hybridization technology. With the aid of the 13 chromosomes nomenclature and the 27 reference BAC markers in sesame, all the 210 BACs are grouped into 13 chromosome groups (data not shown) (Unpublished data, Haiyang Zhang). The relative position of each BAC marker is recorded with more than three replications, and the hybridization signal position in chromosomes is calculated according to the method of Fonsêca et al (2010). The short arm tip is record as 0.00, and the long arm tip is as 1.00. The position value of each BAC ranges between 0.00 and 1.00. Meanwhile, the BAC size is calculated based on positions of the BAC ends sequences in sesame genome.

According to the sesame chromosome nomenclature of Zhao et al. (2018a), *Si*Chr.6 and *Si*Chr.7 possess the 5S rDNA sites in their ends of short arms. The three chromosomes, i.e., *Si*Chr.11, *Si*Chr.12, and *Si*Chr.13 are satellite chromosomes with 45S rDNA sites in the tip of the short arms. In the cytogenetic map, the number of BAC markers in each chromosome varies from 7 to 40 with an average of 16. Further genome location results indicate that the size of the located BACs ranges from 10.7 to 354.6 Kb with an average of 86.0 kb. The reliable chromosome information gives the foundation for sesame genome assembly (Unpublished data, Haiyang Zhang).

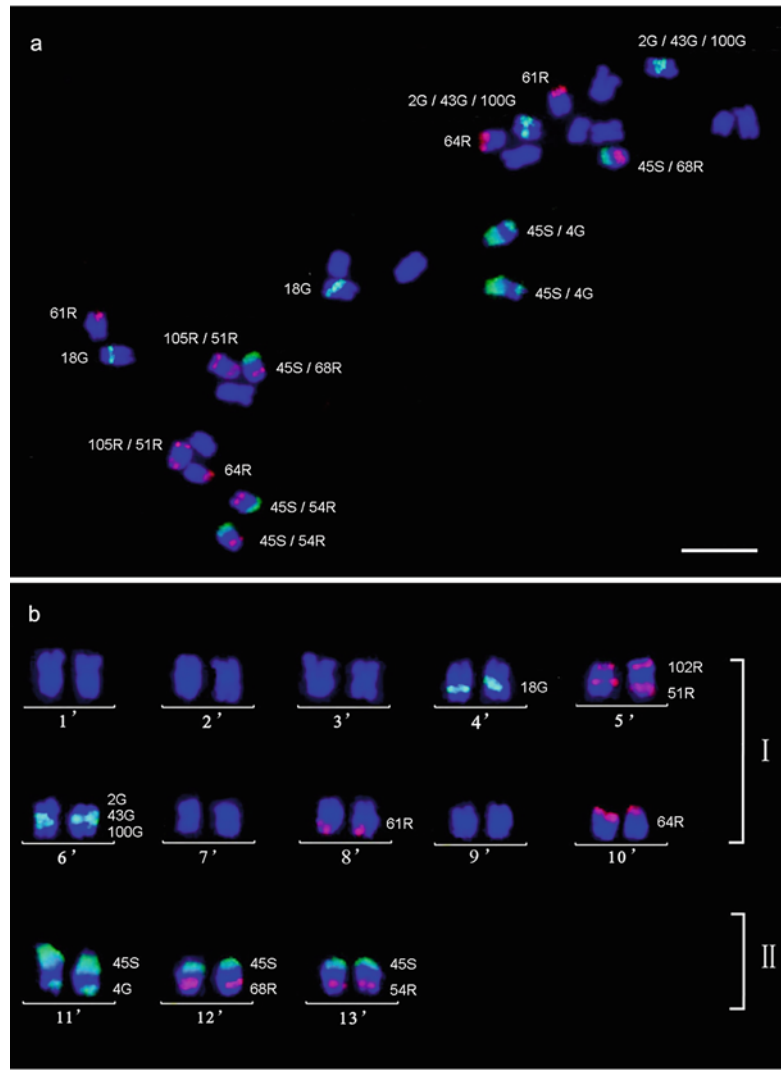


Fig. 9.13 Identification of the three chromosomes with satellite DNAs (Group II). **a** Integration of images from six-round successive fluorescence in situ hybridization. **b** Karyotype analysis of chromosomes hybridized with the 11 BAC probes and 45S rDNA. 'G' and 'R' in probe names refer to green and red fluorescent labeling, respectively. Bar = 5 μm (Modified from Zhao et al. (2018a))

On the other hand, the super-density single nucleotide polymorphism (SNP) genetic map with the 30,193 SNP markers has been constructed for sesame in recent years (Zhang et al. 2016). To evaluate the quality of the cytogenetic map and the SNP genetic map, we aligned the 210 BAC sequences in the cytogenetic map with the 30,193 SNP sites in 13 linkage groups. Results showed that 232 SNPs in the 13 linkage groups are located in the 100 BACs (Unpublished data, Haiyang Zhang). The 13 chromosomes and the 13 linkage groups coordinately accord with each other. Therefore, the BAC-FISH cytogenetic map is reliable and can be used for further genomic analysis. We believe that the integration of cytogenetic map and the molecular genetic map in sesame will display the great effect on the fine assembling of genome and gap filling.

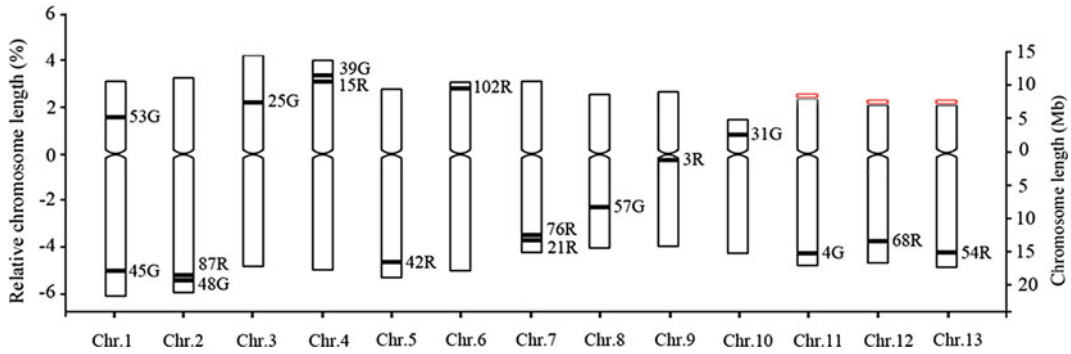


Fig. 9.14 Schematic of the sesame chromosome set marked with 17 BACs. The hollow columns represent the chromosome. Red portions indicate satellite DNA (45S rDNA). Black bars in the hollow columns indicate BAC positions. Vertical coordinates on the left indicate relative chromosomal lengths, while vertical coordinates on the right indicate the genome length of each chromosome, calculated based on the *S. indicum* genome size (Zhang et al. 2013) (Cited from Zhao et al. (2018a))

9.4 GISH and Phylogenetic Analysis in *Sesamum*

9.4.1 GISH of *S. Indicum* and the Wild Species

In the family Pedaliaceae, the basic chromosome number of three genera has been reported, i.e., *Pedalium* ($x = 8$), *Sesamum* ($x = 8$, 13), and *Ceratotheca* ($x = 8$) in 1991 (Kobayashi (1991). However, there exist no other direct proofs to determine the exact basic chromosome number of *Sesamum* except for the above somatic chromosomes analyses (Figs. 9.4, 9.5, and 9.6; Table 9.2). Of the six wild species, *S. alatum* has the same chromosome number with *S. indicum*, while *S. latifolium* has the same rDNA distribution pattern with *S. indicum*. Meanwhile, the four partially cultivated species, i.e., *S. angolense*, *S. calycinum*, *S. angustifolium*, and *S. radiatum* present the species-specific characters in *Sesamum*. In order to explore the cytogenetical evolution and the relationship of the *Sesamum* species, we performed the genome fluorescence in situ hybridization (GISH) analysis within *S. indicum* and the six wild species (Unpublished data, Haiyang Zhang).

Initially, we performed the GISH hybridization of the seven species using the genome of var. Yuzhi 11 as the hybridization probes (Fig. 9.15) (Unpublished data, Haiyang Zhang). The chromosomes of the six wild species are hybridized by genome DNA of var. Yuzhi 11.

For the cultivated sesame, the hybridization signal covers all arms of the somatic chromosomes (Fig. 9.15a). In *S. alatum* (var. 3651, F Fig. 9.15b), *S. angolense* (var. K16, Fig. 9.15d), and *S. calycinum* (var. Ken8, Fig. 9.15e), the hybridization signals possibly present in the repetitive regions such as satellite regions. Meanwhile, in *S. latifolium* (var. Ken1, Fig. 9.15c), *S. angustifolium* (var. G01, Fig. 9.15f), and *S. radiatum* (var. G02, Fig. 9.15g), the signals probably locate in the centromere regions of most chromosomes and the satellite regions of chromosomes. The various distribution styles of the hybridization signals reflect the different similarity of the genome sequences.

9.4.2 GISH of the Wild *Sesamum* Species

In recent a few years, we performed the GISH of the wild *Sesamum* species using genomic DNA of the six wild species as probes (Unpublished data, Haiyang Zhang). The results should be published in near future. Cytogenetical analyses

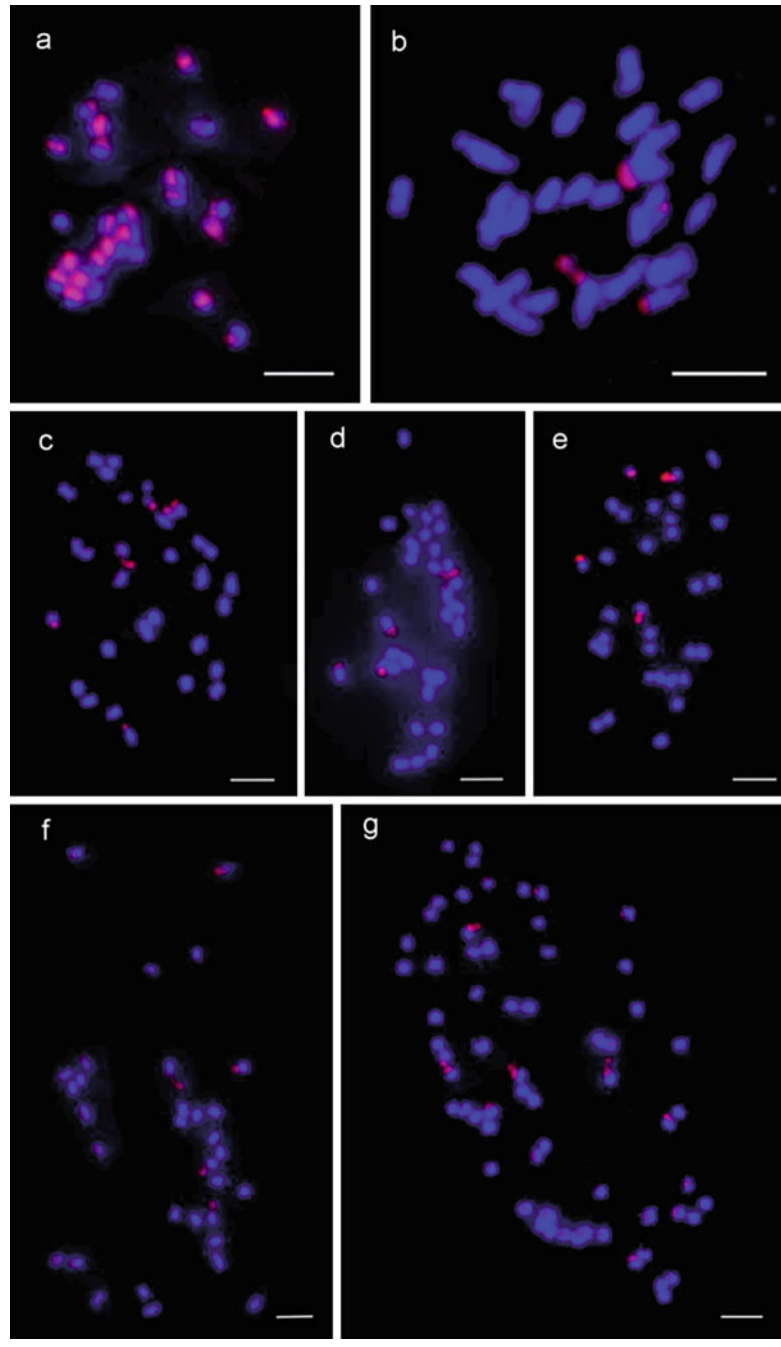


Fig. 9.15 GISH of *S. indicum* in the seven species. **a** Hybridization of *S. indicum* chromosomes (var. Yuzhi 11); **b** Hybridization of *S. alatum* chromosomes (var. 3651); **c** Hybridization of *S. latifolium* chromosomes (var. Ken1); **d** Hybridization of *S. angolense* chromosomes (var. K16); **e** Hybridization of *S. calycinum* chromosomes (var. Ken8); **f** Hybridization of *S. angustifolium* chromosomes (var. G01); **g** Hybridization of *S. radiatum* chromosomes (var. G02). All the chromosome species are hybridized using *S. indicum* (var. Yuzhi 11) probe with red fluorescent labeling. Bar = 5 μm (Provided by Haiyang Zhang)

reflect that some species exhibited the different relationship on chromosome karyotype (Unpublished data, Haiyang Zhang) and supply the basis for further analysis of genome evolution in *Sesamum*.

9.5 Conclusion

The chromosome morphology and the structure of chromosome set in *S. indicum* are explored. The 13 chromosomes in sesame are identified

using 17 BACs for the first time. The cytogenetic map of sesame is constructed based on the high-efficient chromosome press technique and the successive BAC-FISH techniques. To clarify the chromosome evolution of *Sesamum*, the karyotype and chromosome characteristics of the six representative wild *Sesamum* species are investigated. GISH results reflect the relationship and the evolution characters of the chromosome categories ($2n$ = 26, 32, and 64) in *Sesamum*. The results supply the necessary basis for exploring the genome evolution and chromosome evolution in *Sesamum*.

Acknowledgements We would like to thank Dr. Daogang Guan from Qinhui University, Hong Kong, China, for his kindness and guidance for integration of the SNP genetic map and cytogenetic BAC map in sesame.

References

Bedigian D (2015) Systematics and evolution in *Sesamum* L. (Pedaliaceae), part 1: evidence regarding the origin of sesame and its closest relatives. J Plant Taxon Geogr 70:1–42

Benson ON, Beatrice A, Samuel G, Otto GD, Augustino OO (2014) Genetic relationship between sesame (*Sesamum indicum* L.) and related wild species based on chromosome counts and isozyme markers. Afr J Agri Res 9:1052–1060

Chen Y, Li M (1985) Karyotype analyses of four species (varieties) of Lycoris herb. Acta Hort Sin 12:57–60

Chen C, Wang C, Song W (2013a) A Method for fluorescence *in situ* hybridization of 45S rDNA on plant chromosomes (Chinese patent, ZL. 201310346495.9)

Chen C, Wang C, Song W (2013b) A Method for fluorescence *in si*tu hybridization of 5S rDNA on plant chromosomes (Chinese patent, ZL. 201310346485.5)

Fonsêca A, Ferreira J, Santos TRBD, Mosiolek M, Bellucci E et al (2010) Cytogenetic map of common bean (*Phaseolus vulgaris* L.). Chrom Res 18:487–502

He F, He L, Hou L, Liu J (1994) Karyotype as related to phylogenies in sesame. J Southwest Univ Nat Sci Edn 6:573–576

Joshi AB (1961) Sesamum. Indian Central Oilseed Committee Hyderabad, India, p 109

Kobayashi T (1991) Cytogenetics of sesame (*Sesamum indicum*). Dev Plant Genet Breed 2:581–592

Kumar LSS, Abraham A (1941) Acotological study of sterility in *Sesamum orientale* L. Indian J Genet 1:41–60

Liu Y, Cui C, Mei H, Wu K, Zheng Y (2013) Exploration and application of non-denaturing FISH on chromosomes research in sesame (*Sesamum indicum* L.). Sci Agri Sin 46:3729–3735

Liu Y, Cui C, Mei H, Zheng Y (2014) Development and applications of a set of chromosome-specific DNA markers in sesame. Res Crops 15:481–486

Miao H, Zhang H (2014) The sesame genome project and sesame genome sequencing. In: International plant and animal genome conference XXII, Jan 9–13th 2014, San Diego, USA

Morinaga T, Fukushima E (1929) Chromosome numbers of cultivated plants III. Shokubutsugaku Zasshi 45:140–145

Nimmakayala P, Perumal R, Mulpuri S, Reddy UK (2011) Sesamum. In: Kole C (ed) Wild crop relatives: genomic and breeding resources. Volume: Oilseeds. Springer, Berlin, Heidelberg, pp 261–273

Paesold S, Borchardt D, Schmidt T, Dechyeva D (2012) A sugar beet (*Beta vulgaris* L.) reference fish karyotype for chromosome and chromosome-arm identification, integration of genetic linkage groups and analysis of major repeat family distribution. Plant J Cell Mol Biol 72:600–611

Paszko B (2006) A critical review and a new proposal of karyotype asymmetry indices. Plant Syst Evol 258:39–48

Stebbins GL (1971) Chromosomal evolution in higher plants. Edward Arnold (Publishers) Ltd, London, UK

Tran TD, Šimková H, Schmidt R, Doležel J, Schubert I (2017) Chromosome identification for the carnivorous plant *Genlisea margaretae*. Chromosoma 126:389–397

Wang J, Yu J, Sun P, Li Y, Xia R et al (2016) Comparative genomics analysis of rice and pineapple contributes to understand the chromosome number reduction and genome changes in grasses. Front Genet 7:174

Zhan Y, Cheng M, Wu A (1987) Cytogenetical studies in sesame (*Sesamum indicum* L.) I. karyotype analysis. J China Agric Univ 8:249–254

Zhan Y, Cheng M, Wu A (1988) Cytogenetical studies in sesame (*Sesamum indicum* L.) II. Chromosome giemsa banding. J China Agric Univ 7:1–4

Zhan Y, Cheng M, Wu A, Zhou X (1990) Cytogenetical studies in sesame (Sesamum indicum L.). A new taxonomic system. J China Agric Univ 5:11–18

Zhang H, Miao H, Li C, Wei L, Ma Q (2012) Analysis of sesame karyotype and resemblance-near coefficient. Chin Plant Bull 47:602–614

Zhang H, Miao H, Wang L, Qu L, Liu H et al (2013) Genome sequencing of the important oilseed crop *Sesamum indicum* L. Genome Biol 14:401–409

Zhang H, Miao H, Zhao R, Li C, Ma Q et al (2014) A method for the fluorescence in situ hybridization of sesame chromosome (China patent: ZL. 201410322009.4)

Zhang H, Miao H, Li C, Wei L, Duan Y et al (2016) Ultra-dense SNP genetic map construction and identification of *SiDt* gene controlling the determinate growth habit in *Sesamum indicum*L. Sci Rep 6:31556

Zhao R, Miao H, Song W, Chen C, Zhang H (2018) Identification of sesame (*Sesamum indicum* L.) chromosomes using the BAC- FISH system. Plant Biol 20:85–92

Zhao RH, Miao HM, Ma Q, Chen CB, Song WQ et al (2018) Karyotype comparison analysis of the wild species Sesamum alatum and the cultivated sesame. Acta Scientiarum Naturalium Universitatis Nankaiensis 5(05):27–36

Background of the Sesame Genome Project

10

Haiyang Zhang, Lei Wang, and Hongmei Miao

Abstract

With the development of sesame industry in the world, the requirement for increasing the breeding efficiency and realizing various breeding objectives with the aid of modern molecular breeding techniques has become more and more urgent. In order to realize molecular breeding and concisely aggregate the elite genes into new varieties in sesame, deciphering a great amount of genes or molecular markers and elucidating the sesame genome characters are the key issues. Based on the above requisites, Henan Sesame Research Center, Henan Academy of Agricultural Sciences, China organized the Sesame Genome Working Group (SGWG) in 2009 and initiated the Sesame Genome Project (SGP) in 2010. The organizational background of SGWG and the targets of the SGP are also introduced in this chapter.

10.1 Introduction

Traditional breeding techniques including hybridization, mutagenesis, interspecific hybridization, and heterosis are the main methods for sesame breeding in the past forty decades. However, the genetic basis of variation in sesame germplasm is relatively narrow. Screening more elite breeding materials from the limited germplasm accessions becomes too difficult. The vague genetic backgrounds related to the seed yield, seed quality, and the resistances to biotic and abiotic stresses also hinder the development of sesame breeding techniques. In order to elucidate and improve the characters of the sesame species, sequencing the sesame genome, as well as creating more new breeding lines via modern biological techniques, is essential (Zhang et al. 2013).

10.2 The Sesame Genome Working Group

The Sesame Genome Working Group (SGWG) was organized by the Henan Sesame Research Center, Henan Academy of Agricultural Sciences, China (HSRC-HAAS, China) in 2009. The SGWG is composed of the six major sesame research groups in China, which are involved in sesame germplasm investigation and creation, functional genomics and genetics research, biotic and abiotic resistances research, breeding

H. Zhang (✉) · H. Miao
Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou 450002, China
e-mail: zhanghaiyang@zzu.edu.cn

L. Wang
TEDA School of Biological Sciences and Biotechnology, Nankai University, Tianjin 300457, China
e-mail: wanglei@nankai.edu.cn

H. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-98098-0_10

techniques, and/or improvement of processing techniques, in addition to sesame genome sequencing (www.sesamum.org).

10.2.1 The Sesame Genome Working Group of Henan Sesame Research Center, Henan Agricultural Sciences, China (SGWG-HSRC, HAAS, China)

The Henan Sesame Research Center (HSRC), which was sponsored by the Henan Academy of Agricultural Sciences (HAAS), was set up in 1986. HSRC is the leading organization in sesame science research in China and includes 40% of sesame research staff in China. The main scientific research in HSRC focuses on sesame germplasm collection and evaluation, genetics and genomics, breeding technology optimization and selection of new varieties, disease resistance and bio-control techniques, and cultivation techniques research.

The genome research group of HSRC was set up in 2008. The main progress in sesame transcriptomics, molecular marker screening, and genetic map construction was made before initiating the SGP. In order to detect polymorphic molecular markers, a total of 24 groups of sesame transcriptomes were sequenced using Illumina paired-end sequencing technology and a 40G dataset containing 42,566 unitranscript sequences was obtained. More than 3000 expressed sequence tag-simple sequence repeat (EST-SSR) primers were developed using the dataset (Zhang et al. 2012). In order to explore the specific genes involved in sesame growth and development, SGWG-HSRC has constructed a full-length cDNA library of *S. indicum* var. Yuzhi 11 containing 300,000 clones in 2010.

During the SGP initiation, a bacterial artificial chromosome (BAC) (CopyControl™ pCC1BAC™) library of 57,600 clones with an insert size of 85 kb and a BIBAC (pCLD 04,541) library of 80,000 clones with an insert size of 120 kb were constructed by SGWG- HSRC using the cv. Yuzhi 11 for genome assembly and gene cloning (Zhang et al. 2013; Zhao et al. 2018). Fluorescence in situ hybridization (FISH) and BAC- FISH techniques were set up for sesame for the first time in 2013. Subsequently, the successive BAC-FISH technique was developed and utilized for chromosome set differentiation and cytogenetic map construction in sesame. To support the implementation of the Sesame Genome Project, two molecular genetic maps were constructed (Zhang et al. 2012, 2016). Especially, the first high-resolution single nucleotide polymorphism (SNP) genetic map for sesame comprised 13 linkage groups carrying 3010 bin markers and provided the necessary information for super scaffolds assembly. The above necessary platforms and data give the solid foundation for the sesame genome assembly. In addition, the sesame chloroplast genome was published recently (Zhang et al. 2013). The chloroplast genome of *S. indicum* cv. Yuzhi 11 has also been assembled using the raw genome sequencing data, which supplies new proofs for elucidating the phylogenic characters of sesame (Zhang et al. 2012).

In order to accelerate the functional genome research, the technology of sesame callus tissue induction and plantlet regeneration and *Agrobacterium*-mediated genetic transformation were studied by SGWG-HSRC since 2011 (data not shown, Haiyang Zhang). The first transgenic plantlet for sesame was obtained in 2012. At present, the function of some genes is being tested via optimized transgenic methods. Based on the sesame reference genome, SGWG-HSRC established a highly efficient gene cloning method using population mapping and genome variant screening for sesame for the first time (Zhang et al. 2016, 2018; Miao et al. 2019). With the genome information, more than one hundred genes and molecular markers related to seed quality, growth and development, capsule size, and resistance to *Fusarium* wilt disease have been detected in sesame (partially published by Zhang et al. 2016, 2017, 2019; Miao et al. 2019; Wei et al. 2019).

The main goals and tasks of SGWG- HSRC are: (1) to organize and implement the Sesame Genome Project; (2) to construct the genetic

linkage map and physical map of *S. indicum*; (3) to construct transcriptome database; (3) to analyze the genome and chromosome evolution system; (4) to analyze the synthesis and regulation network of sesame fatty acids; and (5) to explore the regulation mechanisms of sesame resistance and tolerance to diseases, waterlogging, and other abiotic stresses.

10.2.2 The Sesame Genome Working Group of Tianjin Biochip Corporation (TBC) and TEDA School of Biological Sciences and Biotechnology, Nankai University (SGWG-TBC, NU, China)

Tianjin Biochip Corporation (TBC) and TEDA School of Biological Sciences and Biotechnology (TSBSB), Nankai University allied and performed the Sesame Genome Project in 2009. The Lab focuses on research and development of microbial detection biochips, genome and transcriptome sequencing and analysis, biological genomics and functional genomics research, and laboratory services. TBC and TSBSB group is one of the five national biochip research and development platforms funded by the National High Technology Research and Development Program of China. The Sesame Genome Working Group of TEDA School of Biological Sciences and Biotechnology (SGWG-TSBSB) was set up in 2009. The chloroplast genome of *S. indicum* cv. Yuzhi 11 was constructed in TSBSB in 2011 (Zhang et al. 2012). During implementation of the Sesame Genome Project, high coverage Solexa sequencing and draft genome assembly platform have been performed in 2011. For sesame, five types of Solexa libraries, including two paired-end libraries with insert sizes of 300 and 500 bp, and three mate-pair libraries with insert sizes of 2, 3, and 5 kb were constructed. A total of 98 Gb of Solexa reads were generated with a 276× coverage of the estimated genome (Zhang et al. 2013). To ensure the assembly quality, the 454 pyrosequencing and BAC end sequencing platforms were set up and performed for sesame during the initiation of the SGP.

The main goals and tasks of SGWG-TSBSB are: (1) to perform the sesame genome sequencing and assembly and (2) to fulfill the sesame BAC library sequencing. During performing the SGP, a complicated sequencing and assembly strategy including ABI3730xL, Roche/454, Illumina/Solexa, PacBio SMRT, BAC-FISH cytogenetic mapping, Bionano mapping, and HiC library platforms were developed and utilized. More data were analyzed accordingly (please see the Chaps. 11 and 12).

10.2.3 The Sesame Genome Working Group of Bioengineering Department, Henan Technology University (SGWG-BD, HTU, China)

The Department of Bioengineering, which is sponsored by the Henan Technology University, focuses on biological and pharmaceutical engineering, biological technology, organism nutriology, and plant pathology. The Sesame Genome Working Group of Bioengineering Department (SGWG-BD) was set up in 2011. The main goals and tasks of SGWG-BD are: (1) to participate in genome sequencing and assembly and (2) to analyze the synthesis and regulation network of sesame fatty acids, proteins, and lignans.

10.2.4 The Sesame Genome Working Group of Plant Protection Research Institute, Henan Academy of Agricultural Sciences (SGWG-PPRI, HAAS, China)

The Sesame Genome Working Group of Plant Protection Research Institute (SGWG-PPRI) was set up in 2009. The research goals are involved screening the germplasm accessions resistant to charcoal rot disease caused by *Macrophomina*

phaseolina, disease resistance genetics and R gene cloning, and functional genomics analysis, and RGAP (resistance gene analog polymorphism) analysis. Four defense enzyme genes (POD, CAT, SOD, and PPO) and eight full-length nucleotide binding site-leucin rich repeat (NBS-LRR) genes have been cloned in SGWG-PPRI Lab before initiating the SGP. The expression characters of some resistance genes induced by *Macrophomina phaseolina* have been analyzed. The main goals and tasks of SGWG-PPRI are: (1) to participate in and perform genome sequencing and assembly and (2) to analyze the regulation mechanism of the resistance and tolerance to charcoal rot disease and waterlogging stress.

10.2.5 The Sesame Genome Working Group of Crop Research Institute, Anhui Academy of Agricultural Sciences (SGWG-CRI, AAAS, China)

The Sesame Genome Working Group of Crop Research Institute, AAAS (SGWG-CRI) was set up in 2009. The research goals involve molecular marker screening and genetic mechanism analysis of male sterility in sesame. The high efficient restoration and maintenance systems about recessive infertility lines have been created by the group during the SGP. The main goals and tasks of SGWG-CRI of AAAS are: (1) to participate in genome sequencing and assembly and (2) to analyze the regulation network of sesame growth and development.

10.2.6 The Sesame Genome Working Group of Crops Research Institute, Jiangxi Academy of Agricultural Sciences (SGWG-CRI, JAAS)

The Sesame Genome Working Group of Crops Research Institute, JAAS (SGWG-CRI) engages in the genetics analysis of seed quality and resistance and tolerance to biotic and abiotic stresses, black sesame breeding, and seed multiplication in sesame. The group also performs molecular markers and genes (or QTLs) detection related to the key agronomic traits, such as inflorescence habit, seed yield, seed quality, and stress tolerance. The main goals and tasks of SGWG-CRI of JAAS are: (1) to participate in genome sequencing and assembly and (2) analyze the regulation network of sesame resistance and tolerance to bacterial diseases and drought stress.

10.3 The Sesame Genome Project

The development of the next-generation high-throughput sequencing techniques has highly impacted the breeding techniques in crops. Considering the status and key hurdles of sesame breeding techniques, the Chinese sesame scientist in chief, Dr. Haiyang Zhang, conceived of sequencing the sesame genome in 2008. Subsequently, Dr. Haiyang Zhang called together the main sesame research groups in China and organized the Sesame Genome Working Group (SGWG) in 2009. The blueprint for the Sesame Genome Project (SGP) was then determined by the SGWG in 2010 (Zhang et al. 2013). According to the original design, the project comprised three parts: (1) to sequence the sesame genome and construct the frame and the fine genome map, respectively; (2) to perform the sesame genome informatics analysis; and (3) to perform the sesame functional genomics research and database construction. The total genome goals of the sesame genome would be completed by December 2013. At that time, the feasible sequencing platform was the next-generation sequencing (NGS) (such as Illumina sequencing) with de novo genome assembly technique (Mitchelson 2007; Shulaev et al. 2011).

However, the unexpected fast development of genome sequencing and assembly techniques during 2010–2015 enriched the sesame genome project. Especially, single-molecule real-time (SMRT) sequencing (Rhoads and Au 2015),

optimal mapping (such Bionano) (Shelton et al. 2015), and HiC sequencing (Servant et al. 2015) were gradually applied in sesame genome sequencing in the past five years and achieved the assembly of the chromosome-scale fine sesame genome. Correspondingly, the SGP expanded with four new targets: (1) to construct the chromosome-scaled genomes of sesame and several wild *Sesamum* species and the *Sesamum* genome databases; (2) to perform the comparative genomics analysis of *Sesamum* species; (3) to perform genome structure and evolution analysis of *Sesamum* species; and (4) to perform genome re-sequencing of sesame germpalsm accessions and genome application in the molecular genetics and breeding research in sesame. Totally, the SGP went through 10 years, as the SGWG hit the above research targets and declared the fulfilling of the SGP on June, 8, 2020 (Unpublished data, H. Zhang). The main research results of the SGP will be gradually published in near future.

10.4 Timeline of the Sesame Genome Project

According to the above original blueprint for the SGP, all the goals would be completed in 2013. Of these, the SGWG would finish the 454 paired-ends reads sequencing by December 2012, and finish the double-ended sequencing of the 40,000 BAC clones and full-length sequencing of the 1000 BAC clones by June, 2013. Independently, a high efficient sequencing platform, PacBio SMRT technique had been developed and was being applied for genome sequencing and assembly because of the long length of the reads at that time. Thus, the SGWG optimized the sequencing strategy of the SGP and applied the PacBio SMRT in sesame genome sequencing in 2013. Subsequently, the workflow of the genome assembly using ABI3730 xL, Solexa, Roche/454, and PacBio SMRT was rapidly optimized for sesame. The SGWG completed the analysis result and the de novo assembly of sesame genome using Solexa/Illumina data in 2012 and expanded the goals of the SGP in 2013 (Zhang et al. 2013, 2019). During the past decade, the genome assembly version of sesame has been updated three times (Zhang et al. 2013, 2016; Miao et al. 2019), as more new sequencing and assembly techniques, were gradually applied in the SGP.

10.5 Biological Questions and Planned Analyses of Sesame

According to the initiation of SGP, the SGWG planned to elucidate several key biological questions including the phylogeny and the origin of sesame, the biosynthesis and regulation of oil, protein, and some secondary metabolites, the resistance to *Fusarium* wilt, charcoal rot diseases, and waterlogging stress (Zhang et al. 2013). With the expansion of the goals of SGP, the genomes of several wild *Sesamum* species were sequenced and assembled. Phylogenic and genome evolution characters of the *Sesamum* species have been systematically analyzed. The genomic mechanism to the low resistance to diseases and high oil content in sesame also has been studied (Unpublished data, Haiyang Zhang). With the aid of the reference genome data and the genome analysis results, several sesame genes related to the inflorescence determinacy, internode length, capsule indehiscence, capsule size, and other agronomic traits have been cloned and published or being published in a series of publications (Zhang et al. 2016, 2018, 2019; Miao et al. 2019).

10.6 Status of Current Genome Assemblies

According to the estimation of the flow cytometry data, the genome size of sesame is about 369 Mb (Zhang et al. 2013). The first draft genome version sequenced by Solexa platform for sesame cv. Yuzhi 11 is 293.7 Mb in length (Zhang et al. 2013). A total of 23,713 gene models were obtained and the average gene length was 1.2 kb. During the second five years,

the SGWG sequenced and assembled the six wild *Sesamum* species covering all the three chromosome sets, i.e., $2n = 26$ (*S. alatum*), $2n = 32$ (*S. latifolium*, *S. angolense*, *S. calycinum*, and *S. angustifolium*), and $2n = 64$ (*S. radiatum*). The genome of *S. indicum* (version 3.0) has also been reassembled using new sequencing and assembly strategy. All the genomes of sesame and the six wild species reach to chromosome-scale and have been applied for genomics and comparative genomics analysis. Some genomics analysis results of the *Sesamum* species will be shown in the following chapters.

References

Miao H, Li C, Duan Y, Wei L, Ju M et al (2019) Identification of a *Sidwf1* gene controlling short internode length trait in the sesame dwarf mutant *dw607*. Theor Appl Genet 133(1):73–86

Mitchelson KR (2007) New high throughput technologies for DNA sequencing and genomics. Science press, Beijing, China

Rhoads A, Au KF (2015) PacBio sequencing and its applications. Genom Proteom Bioinformat 13(5):278–289

Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C et al (2015) HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol 16 (1):259

Shelton JM, Coleman MC, Herndon N, Lu N, Lam ET et al (2015) Tools and pipelines for BioNano data: molecule assembly pipeline and FASTA super scaffolding tool. BMC Genom 16:734

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O et al (2011) The genome of woodland strawberry (*Fragaria vesca*). Nat Genet 43(2):109

Wei L, Li C, Duan Y, Qu W, Wang H et al (2019) A SNP mutation of *SiCRC r*egulates seed number per capsule and capsule length of cs1 mutant in sesame. Intl J Mol Sci 20(16):4056

Zhang H, Miao H, Li C, Wei L, Ma Q (2012) Analysis of sesame karyotype and resemblance-near coefficient. Chin Bull Bot 47(6):602–614

Zhang H, Miao H, Wang L, Qu L, Liu H et al (2013) Genome sequencing of the important oilseed crop *Sesamum indicum* L. Genome Biol 14(1):401

Zhang H, Miao H, Li C, Wei L, Duan Y et al (2016) Ultra-dense SNP genetic map construction and identification of *SiDt* gene controlling the determinate growth habit in *Sesamum indicum* L. Sci Rep 6:31556

Zhang H, Miao H, Li C, Wei L, Duan Y (2017) Sidt1 gene controlling determinate growth habit in sesame and SNP molecular marker thereof. Patent US20170233831A1

Zhang H, Miao H, Wei L, Li C, Duan Y et al (2018) Identification of a *SiCL1* gene controlling leaf curling and capsule indehiscence in sesame via cross-population association mapping and genomic variants screening. BMC Plant Biol 18(1):296

Zhang H, Miao H, Ju M (2019) Potential for adaptation to climate change through genomic breeding in sesame. In: Kole C (ed) Genomic designing of climate-smart oilseed crops. Springer, Cham, Switzerland, pp 374–376

Zhao R, Miao H, Song W, Chen C, Zhang H (2018) Identification of sesame (*Sesamum indicum* L.) chromosomes using the BAC-FISH system. Plant Biol 20:85–92

Sequencing of Chloroplast Genome in Sesame

11

Chun Li, Haiyang Zhang, Ming Ju, Fangfang Xu, and Hongmei Miao

Abstract

In plants, chloroplast genomes are a group of non-nuclear genomes and exhibit different genome structure and evolutionary characters from the nuclear genomes. In order to investigate the species-specific evolutionary characters, the complete chloroplast (cp) genome (GenBank no. JN637766) of *S. indicum* cv. Ansanggae (black-seeded) and the cp genome (GenBank no. KC569603) of *S. indicum* cv. Yuzhi 11 (white seeded) have been sequenced. Comparison of the cp genomes between sesame and other plants and the application of the cp genome information in genome evolution analysis are elucidated in this chapter.

11.1 Introduction

The chloroplast is a vital plastid existing in plants and algae and takes the important functions in plant photosynthesis and related biological processes (Sugiura 2003). Chloroplast genome is an independent and unnuclear genome in plant cell. As the chloroplast genome sequence of a species is relative conservative, the chloroplast genome is always used for species evolution analysis and genetic modification and breeding in crops (Corriveau and Coleman 1988; Sugiura 1989; Säll et al. 2003; Ruf et al. 2007; Saski et al. 2011; Young et al. 2011). In plants, the circular cp genome size is usually 120–160 kb with about 4 rRNAs, 30 tRNAs, and 80 genes encoding the photosynthesis or gene expression regulation-related proteins (Olmstead and Palmer 1994; Chumley et al. 2006). To date, more than one hundred cp genomes have been sequenced (Chloroplast Genome DB, http://chloroplast.cbio.psu.edu).

Sesame belongs to the Pedaliaceae family which is a small family. Thus, sesame is always regarded as an orphan species for the remote phylogenetic relationship with other common crops (Ashri 1998). Previous genome phylogenetic position analysis of so many plants from different families indicated that *Sesamum* is closely related to members of the Solanaceae and Phrymaceae families, but distant to the other oil crops such as soybean, castor, and rape (Zhang et al. 2013b).With the aid of the chloroplast (cp) genome sequence of *S. indicum* cv. Ansanggae (a black-seeded cultivar), Yi and Kim (2012) performed the phylogenetic position analysis of sesame and suggested that *Sesamum* is a sister genus to the *Olea* and *Jasminum* (Oleaceae family) and belongs to the core lineage of the Lamiales family. During sequencing the genome of cv. Yuzhi 11 in the Sesame Genome

C. Li · H. Zhang (✉) · M. Ju · F. Xu · H. Miao
Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, People's Republic of China
e-mail: zhanghaiyang@zzu.edu.cn

H. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-98098-0_11

Project (SGP), the chloroplast genome of *S. indicum* was also assembled using Illumina and 454 sequencing data and applied for species evolution analysis (Zhang et al. 2013a). As to the mitochondria genome in sesame, Li et al. (2011) studied the extraction technology of mitochondrial DNA in sesame. However, no mitochondria genome information about sesame has been reported till now. Thus, we introduce the chloroplast genomes of cv. Yuzhi 11 and cv. Ansanggae, respectively, and exhibit the main analysis results of the cp genome comparison of *Sesamum* species with other plants in this section.

11.2 Sequencing of the Complete Chloroplast Genome in Sesame *cv.* Ansanggae

Yi and Kim (2012) performed sequencing and assembly of the complete chloroplast genome for sesame for the first time. Cp organelles of the sesame cultivar cv. Ansanggae (a black-seeded cultivar) plant were collected for Cp DNAs extraction (Plant DNA Bank of Korea accession number 1996–0001). The chloroplast DNA sequences were sequenced using the GS-FLX pyrosequencing method (Wyman et al. 2004) and the Genome Sequencer FLX system (Roche, Basal, Switzerland). A total of 133,533 reads of Roche 454 data with an average read length of 236 bp were obtained for genome assembly. As a result, the complete cp sequence of *S. indicum* cv. Ansanggae is 153,324 bp in length (GenBank accession no. JN637766). A total of 114 genes including 80 protein-coding genes, 30 tRNA genes, and 4 rRNA genes are detected in the cp genome.

11.3 Sequencing of the Complete Chloroplast Genome in Sesame *var.* Yuzhi 11

The complete cp genome of *S. indicum* var. Yuzhi 11 was assembled using Illumina and 454 sequencing data of sesame using de novo assembly strategy (Zhang et al. 2013a). For sesame genome sequencing, paired-end (PE) and mate-pair (MP) libraries with insert sizes of 500 bp and 3 Kb, respectively, were constructed and sequenced on Illumina sequencing platform. The Illumina data gave a coverage of 218× the estimated cp genome. Meanwhile, 4.5 Gb of Roche 454 data were obtained from GS FLX sequencing system (Zhang et al. 2013a). In order to assemble the chloroplast genome, all quality-filtered paired reads were mapped against the cp genomes of *Ageratina adenophora* (NCBI: NC_015621.1) and *Olea europaea* (NCBI: NC_015623) using BWA (v 0.5.8) for the cp genome assembly for sesame (Zhang et al. 2013a). As a result, the assembled cp genome of *S. indicum var.* Yuzhi 11 is 153,338 bp with 114 unique genes (GenBank accession no. KC569603) (Fig. 11.1). The circular molecular of the cp genome contains three scaffolds and is comprised of the inverted repeat (IR, 25,142 bp), large single-copy (LSC, 85,180 bp), and small single-copy (SSC, 17,874 bp) regions.

11.4 Characters of Complete Chloroplast Genome in Sesame

Comparison of the cp genomes of var. Yuzhi 11 and cv. Ansanggae showed that only 14 differences existed within the nucleotide sequences of homopolymers (Zhang et al. 2013a). The results reflect the conservation of chloroplast genome in sesame. Compared with the 18 species presenting the available nuclear genome sequences, the cp gene number in sesame is also same with those of *Nicotiana tabacum*, *Vitis vinifera*, *Platanus occidentali*. However, in some plants such as *Arabidopsis thaliana*, *Glycine max*, *Brassica napus* and *Mangifera indica*, one or two genes were lost. Meanwhile, the order of some genes in the sesame cp genome is different from that of *G. max*, *Helianthus annuus* and *Gossypium hirsutum*.

To further clarify the evolutionary position of sesame in angiosperms, Zhang et al. (2013a) performed the phylogenetic analysis of

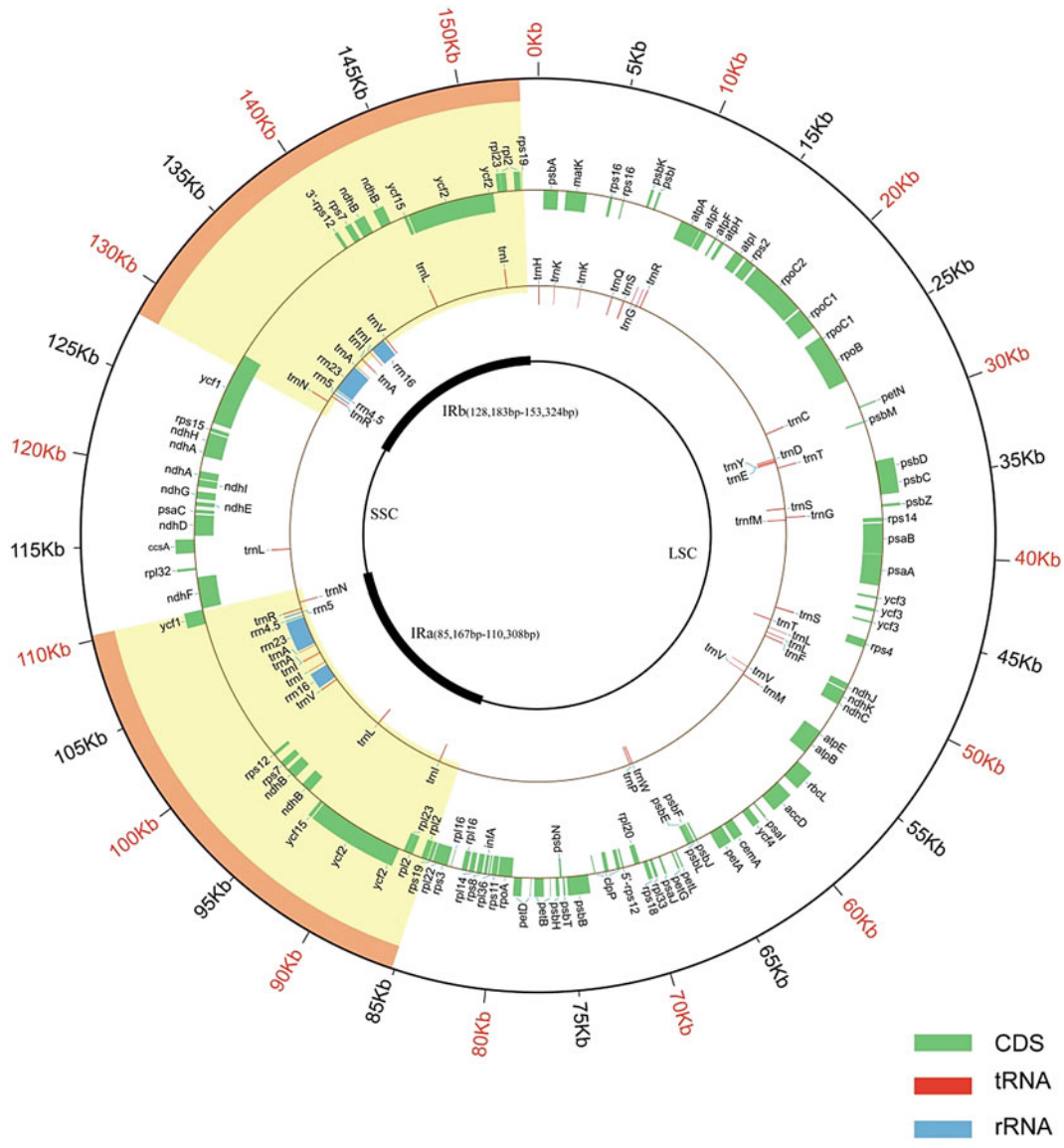


Fig. 11.1 Chloroplast genome map of sesame var. Yuzhi 11. The outer line indicates the chloroplast genome size. The middle two circle lines represent the position of all protein-coding genes, rRNAs, and rRNAs the IRa and IRb inverted repeat sequences, respectively, in clockwise direction. The inner circle indicates distribution of SSR, LSC, and IR regions (Cited from Zhang et al. (2013a))

chloroplast genomes of sesame and the other 18 plants. The results showed that the cp size of sesame is moderate. Variation of the length of LSC and IRs regions contributed to the change of a cp genome size. Moreover, Zhang et al. (2013a) compared the protein-coding sequences of the 19 cp genomes for the first time. Of the 80 protein-coding genes, the length of all 77 functional cp genes was highly conserved. The *ycf2* gene varied in some plants and even was missing in some grass species. Interestingly, in sesame cp genome, the *ycf2* gene is 5721 bp in length with a loss of a fragment of about 1179 bp. BLAST results proved that nucleotides 1–585 in the

query hit the sesame nuclear genome, which might reflect the specific evolution relationship between the nuclear and unnuclear genomes in high plants.

11.5 Evolution Analysis of Chloroplast Genome

Comparison results indicated that IR contraction/expansion in the cp genome alters the evolution rate of cp genome in plants. Meanwhile, the chloroplast genes and repeats also displayed the signature of convergent evolution in sesame and other species (Zhang et al. 2013a). The *Ka*/*Ks* ratio of the 77 protein-coding genes in sesame and 13 other dicot species from the asterid and rosid clades indicated that the evolutionary rates of cp genes were different. In sesame cp genome, nine genes, i.e., *ndhB*, *ndhD*, and *ndhI* genes encoding the subunits of NADH dehydrogenase, the *rpl2*, *rpl22*, *rpl32*, and *rpl33* genes encoding the large subunit of the ribosome, the *rps12* gene encoding the small subunit of the ribosome, and the *rbcL* gene encoding the large subunit of Rubisco, evolved rapidly probably because of the specific function. Three genes, i.e., *ndhK* encoding the subunit of NADH dehydrogenase, *atpI* encoding the subunit of ATP synthase, and *cemA* encoding the envelope membrane protein evolved slowly with low evolutionary rates.

Recently, the Sesame Genome Working Group (SGWG) also assembled the six chloroplast genomes of the six wild *Sesamum* species using the abundant genome sequencing data. The genome analysis is being performed. We believe the new findings of the evolutionary characters of *Sesamum* species and the higher plants would improve the genome research of *Sesamum* in near future.

References

Ashri A (1998) Sesame breeding. In: Janick J (ed) Plant breeding reviews, vol 16. Wiley, Israel, pp 179–228

Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ et al (2006) The complete chloroplast genome sequence of *Pelargonium* x*hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol 23:2175–2190

Corriveau JL, Coleman AW (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperms. Amer J Bot 75:1443–1458

Li D, Wang L, Zhang Y, Lu H, Zhang X et al (2011) The study on the extraction technology for pure mitochondrial DNA from sesame. Acta Agri Bor Sin 3:90–94

Olmstead RG, Palmer JD (1994) Chloroplast DNA systematic: a review of methods and data analysis. Amer J Bot 81:1205–1224

Ruf S, Karcher D, Bock R (2007) Determining the transgene containment level provided by chloroplast transformation. Proc Natl Acad Sci USA 104:6998–7002

Säll T, Jakobsson M, Lind-Halldén C, Halldén C (2003) Chloroplast DNA indicates a single origin of the allotetraploid *Arabidopsis suecica*. J Evol Biol 16:1019–1029

Saski C, Lee SB, Fjellheim S, Guda C, Jansen RK et al (2011) Chloroplast genome variation in upland and lowland switchgrass. PLoS ONE 6:e23980

Sugiura M (1989) The chloroplast chromosomes in land plants. Annu Rev Cell Biol 5:51–70

Sugiura M (2003) History of chloroplast genomics. Photosynth Res 76:371–377

Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252–3255

Yi DK, Kim KJ (2012) Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. PLoS ONE 7:e35872

Young HA, Lanzatella CL, Sarath G, Tobias CM (2011) Chloroplast genome variation in upland and lowland switchgrass. PLoS ONE 6:e23980

Zhang H, Li C, Miao H, Xiong S (2013a) Insights from the complete chloroplast genome into the evolution of *Sesamum indicum* L. PLoS ONE 8(11):e80508

Zhang H, Miao H, Wang L, Qu L, Liu H et al (2013b) Genome sequencing of the important oilseed crop *Sesamum indicum* L. Genome Biol 14:401

Strategies and Tools for Sequencing of the Sesame Genome

12

Hongmei Miao, Yamin Sun, Lei Wang, and Haiyang Zhang

Abstract

The invention of Sanger sequencing initiated the genome era. The development and application of next-generation sequencing (NGS) technologies facilitated genome sequencing in organisms. Here we introduce the main technologies including Sanger sequencing, Roche-454 massive parallel pyrosequencing, Solexa/Illumina sequencing, ABI SOLiD platform, PacBio SMRT platform, HiC sequencing, Oxford nanopore sequencing, and Bionano sequencing platform, as mostly are applied in the Sesame Genome Project (SGP). The characteristics of each technology are briefly discussed. For the SGP, the hybrid sequencing strategies involving the first generation, NGS, and the third generation sequencing technologies are applied to assemble the fine genome map with high accuracy and completeness. Combined with the genome size estimation and genome characters of sesame, the efficient sequencing strategies are elucidated in this section.

12.1 Introduction

The invention of Sanger sequencing initiated the genome era in the early of 1990s. The earliest sequenced genomes using Sanger sequencing are *E. coli* genome (Yura et al. 1992) and *C. elegans* genome (Sulston et al. 1992). In plants, the genome of the model plant *Arabidopsis* was sequenced using BAC by BAC strategy in 2000, which is regarded as a milestone for plant research (Bolger et al. 2014). In 2006, the whole genome shortgun sequencing strategy was applied for sequencing the genome of poplar (*Populus tirchocarpa*) for the first time. The genomic DNA was randomly broken into small pieces for Sanger sequencing. As a result, approximately 7.6 million end-reads were sequenced and assembled into 2447 major scaffolds (Tuskan et al. 2006). Even though the whole genome shotgun sequencing strategy was being applied in plant genome research, the genome sequencing technologies were still time consuming with high cost during the early 2000s.

The invention of the next-generation sequencing (NGS) technologies accelerated the genomics research (Mitchelson 2007). Combined with the Sanger sequencing, the next-generation Illumina GA sequencing technologies were successfully applied for the cucumber (*Cucumis sativus*)

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

Y. Sun · L. Wang
TEDA School of Biological Sciences and Biotechnology, Nankai University, Tianjin, P. R. China
e-mail: nksunyamin@aliyun.com

L. Wang
e-mail: wanglei@nankai.edu.cn

H. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-98098-0_12

genome in 2009. The genome coverage of sequencing data reached to 72.2-fold (Huang et al. 2009). Due to the small genome size (240 Mb), the woodland strawberry (*Fragaria vesca*) becomes the first crop sequenced independently using the NGS and the *de nove* assembly strategy (Shulaev et al. 2011).

The NGS technologies offer advantages for genome sequencing over Sanger sequencing. However, the defect of the short read length limits the application of NGS in the assembly of large and complex genomes. Single-molecule real-time (SMRT) sequencing developed by Pacific BioSciences (PacBio) offers long read length and overcomes the hurdles of the NGS technologies in genome assembly (Rhoads and Au 2015). The early assembled plant genomes using SMRT sequencing data alone are *Arabidopsis* (Berlin et al. 2015) and *Oropetium thomaeum* (Vanburen et al. 2015). A total of 99% (244 Mb) of the *Oropetium thomaeum* genome was assembled into the 625 contigs with an N50 length of 2.4 Mb. The *de novo* assembly using PacBio sequencing supplies a new and feasible approach for genome sequencing and assembly. Subsequently, the Nanopore sequencing independently applied in bacterial genome research also gives the example for accurate genome construction (Loman et al. 2015). Throughout the history of the sequencing technology development, each generation of the sequencing technologies and platform reflects the development of the sequencing technology and the progress of genome research (Fig. 12.1).

Recently, the 10KP (10,000 Plants) Genome Sequencing Project covering embryophytes, green algae, and protists (excluding fungi) was put forward and would be performed within the next five years (Cheng et al. 2018). Correspondingly, the optimized sequencing methodologies and strategies for specific species should be reasonably designed, in order to improve the sequencing quality and to decrease the expenses. In this chapter, we discuss on the main sequencing technologies including the first generation, NGS, and the third generation sequencing platforms. The combined sequencing strategy for sesame genome is concisely described.

12.2 Genome Sequencing Technologies

12.2.1 Sanger Sequencing

The Sanger sequencing method is also named 'first-generation' sequencing technology, and is a chain termination method. Sanger sequencing was reported by Fred Sanger and his colleagues in 1977 (Sanger et al. 1977). In Sanger sequencing, 500–1000 base pairs (bp) of the high-quality DNA sequences in length are routinely sequenced using Applied Biosystems.

Sanger sequencing involves a series of processes, i.e., bacterial cloning or PCR (polymerase chain reaction) preparation, template purification, DNA fragments labeling using the chain termination method with dye-lablelled dideoxynucleotides, capillary electrophoresis, and fluorescence detection of four-color plots to reveal the DNA sequence. Sanger sequencing produces many copies of a target DNA region through PCR. In the reaction, dideoxynucleotides are similar to deoxynucleotides, except for lacking a hydroxyl group on the 3' carbon of the sugar ring. The chain ends with the dideoxynucleotide, which is marked with a particular color of dye depending on the base type (A, T, C, or G). In the capillary gel electrophoresis process, the DNA fragments in reactions run through a long and thin tube containing a gel matrix. As each fragment crosses the 'finish line' at the end of the tube, the products are illuminated by a laser, and dye bases are detected. The smallest fragment (ending just one nucleotide after the primer) crosses the finish line first, followed by the longer fragments (ending two and more nucleotides gradually after the primer). Based on the detected dye colors one after another on the detector, a series of peaks in fluorescence intensity are recorded. Then the DNA sequence is read from the peaks in the chromatogram.

The Sanger sequencing machines produce the relatively long reads of slightly less than 1 kb. In order to sequence longer fragments and mega-genomes, researchers developed the shortgun DNA sequencing strategy based on Sanger

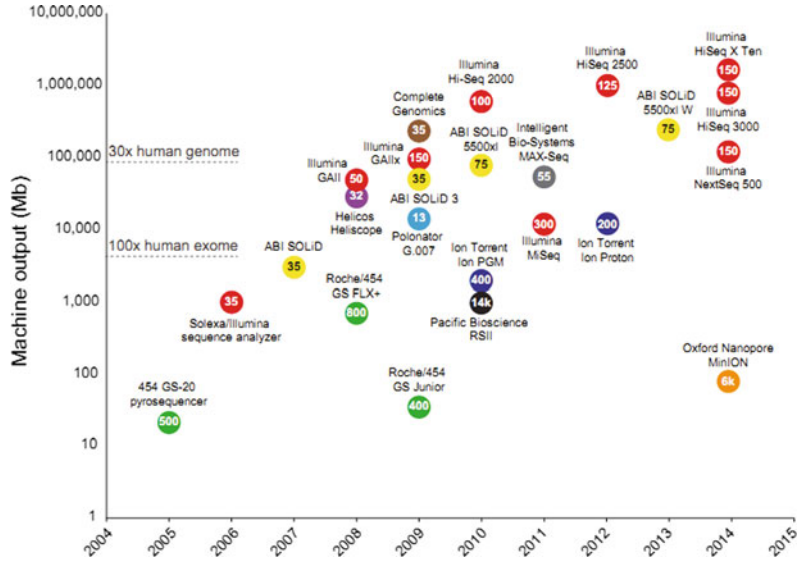


Fig. 12.1 Timeline and comparison of the commercial high throughput sequencing instruments. (Modified from Reuter et al. (2015))

sequencing approach in 1970s (Anderson 1981). The overlapped sub-fragments could be cloned and sequenced separately and then assembled into one long contiguous sequence (named 'contig') in silico (Heather and Chain 2016). In the Human Genome Project (HGP), the modified whole shortgun sequencing approach was applied. Different plasmid libraries with 2 kb, 10 kb, and 50 kb insert sizes, respectively were constructed. The small fragments (500–750 bp) of human DNA were sequenced using Sanger sequencing with ABI PRISM 3700 DNA Analyzer (Venter et al. 2001). The average trimmed sequence length was 543 bp.

Sanger sequencing supplies high-quality sequences for relatively long DNA fragments (up to about 900 bp). Thus Sanger sequencing and the modifications keep the dominate position in DNA sequencing field for nearly 30 year. However, Sanger sequencing is an expensive and inefficient method, if applied for large-scale genome projects. To date, the technology is mainly and typically used for sequencing short DNA sequences, such as bacterial plasmids or PCR products.

12.2.2 Roche-454 Sequencing

Differing from the first generation sequencing, the NGS sequencing technologies can generate huge genome-wide data and deeply impact the biological and genetic research (Metzker 2010; Das and Tan 2013; Heather and Chain 2016). The NGS sequencing technologies such as Roche-454 massive parallel pyrosequencing apply a combination of template preparation, sequencing and imaging, and genome alignment and assembly methods (Metzker 2010). Roche-454 massive parallel pyrosequencing platform sequencing or GS FLX sequencing system was developed by 454 Life Sciences (later purchased by Roche) in 2005 and became the first NGS technology. In Roche 454 sequencing platform, DNA fragments are isolated and amplified using emulsion-based PCR method (Voelkerding et al. 2009). The pyrophosphate-based sequencing is applied (Margulies et al. 2005).

As for DNA library construction in 454 sequencing system, the entire genomic DNA is sheared into small fragments of 300–800 bp using spray method. The fragments are ligated to

the specialized common adapters, separated into single strands, and fixed to 28 μm beads. The beads are captured within the droplets of a PCR-reaction-mixture-in-oil emulsion.

The key feature of emulsion PCR (emPCR) lies in the formation of a large number of independent reaction spaces for DNA amplification. PCR reaction occurs in each droplet. Ideally, each small drop of liquid contains only one DNA template and one bead. Each bead can carry ten million copies of a unique DNA template. As the emulsion is broken, the DNA strands are denatured, and beads carrying single-stranded DNA clones are deposited into wells of a fibre-optic slide (Margulies et al. 2005). Each fibre-optic core is 44 μm in diameter and surrounded by 2–3 μm of cladding. The reaction well is about 55 μm in depth, and the well density is 1.6 million wells (or 480 wells per mm^2) in a slide (Leamon et al. 2003).

In the modified pyrosequencing, DNA sequencing reaction is performed based on the single stranded DNAs. If one dNTP matches with the template DNA, the pyrophosphate group will be released after synthesis. The released pyrophosphate group reacts with ATP sulfurylase enzyme and produces ATP. CO-oxidation of ATP and luciferase produces light that is proportional to the amount of pyrophosphate. The charged couple device (CCD) sensor on the other side of the PicoTiterPlate (PTP) board captures the emitted photons of the fluorescent (Nyrén and Lundin 1985; Leamon et al. 2003). As each kind of dNTP produces the unique fluorescence color in the reaction, DNA sequence can be determined according to the various fluorescence colors. After amplification, ATP is degraded by diphosphatase and leads to fluorescence quenching. Thus the sequencing reaction goes into the next cycle. The obvious advantage of 454 sequencing technology is to get long sequencing read length up to 400 bp. The limitation of the Roche-454 technology relates to the homopolymers. In particular, during sequencing the homopolymers with increasing length, the insertion and deletion errors are often introduced into the sequencing results (Huse et al. 2007).

12.2.3 Illumina Solexa Sequencing

For NGS sequencing technologies, the unique combination of specific protocols supplies the distinction between the technologies and platforms (Metzker 2010). The Solexa sequencing method (later acquired by Illumina) is the most important NGS technology. Of all the available NGS technologies, the Illumina data, especially generated on the HiSeq, gives the highest quality base calls. To date, the Illumina Solexa sequencing platform still occupies a vast part of the NGS market.

Solexa Genome Analyzer (Illumina, San Diego, USA) was commercialized in 2006 is the first short read sequencing technology (https://www.illumina.com). Illumina sequencing technology uses clonal amplification and sequencing by synthesis (SBS) chemistry to perform the rapid and accurate sequencing (https://www.illumina.com). The Solexa/Illumina sequencing consists of three processes, i.e., library preparation, cluster amplification, and post-sequencing data processing (Fig. 12.2). A flow cell consisting of an optically transparent slide with 8 individual lanes is used.

During library preparation, the template DNA is sheared into short fragments with several hundred base pairs in length. After the ends of the fragments are repaired, single adenine (A) bases are added to the 3' ends of the blunt phosphorylated fragments. Then adapters are ligated to the DNA fragments. Adapted-modified DNA fragments are size screened and amplified to improve the quality of templates. Before sequencing, the single-stranded DNA fragments are attached to a flow cell via hybridization. The free 3' end of the adapters is then bonded to a neighboring complementary oligonucleotide to the flow-well anchors. A bridged structure forms. Then the solid phase amplification is performed in the flow cell by cluster PCR (dubbed 'bridge PCR') through the addition of unlabeled nucleotides and other necessary reagents (Shendure and Ji 2008). In this platform, each cluster consists of about 1000 clonal amplicons. Several millions of clusters can be amplified in each eight

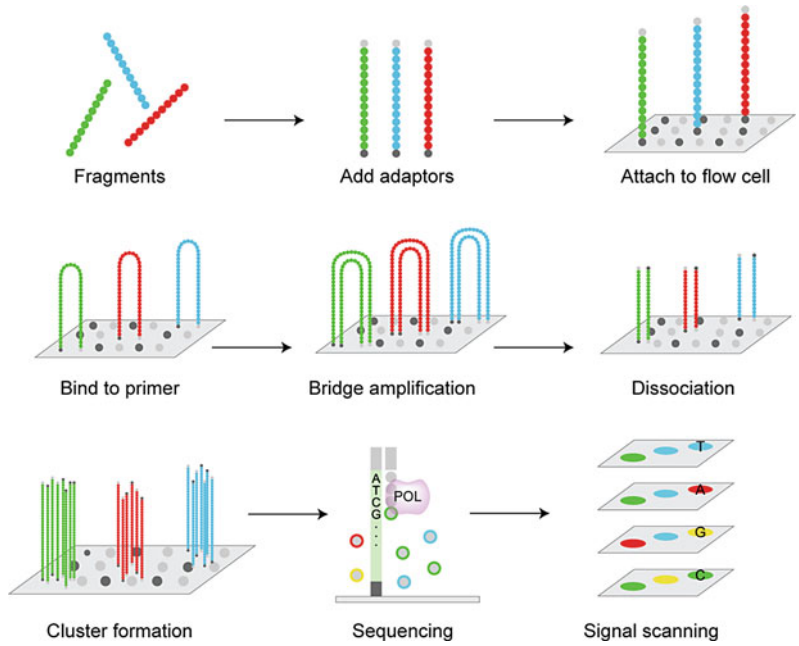


Fig. 12.2 Illumina genome analyzer sequencing. Adapter-modified, single-stranded DNA is added to the flow cell and immobilized by hybridization. Bridge amplification generates clonally amplified clusters. Clusters are denatured and cleaved; sequencing is initiated with addition of primer polymerase (POL) and 4 reversible dye terminators. Postincorporation fluorescence is recorded. The flour and block are removed before the next synthesis cycle (Cited from Voelkerding et al. (2009))

lanes on a single flow-cell (Shendure and Ji 2008). The sequencing reactions occur within each cluster, with the following steps including the incorporation of fluorescent-labeled terminator nucleotides (chemically blocked 3' hydroxyl ends), excitation, and image acquisition. This cycle is repeated for each nucleotide of the sequence (Shendure and Ji 2008) (https://www.illumina.com).

With the development of the techniques, Illumina/Solexa platform comprises of multiple instruments with various throughput data in different read length. The MiniSeq and MiSeq instruments offer low to mid sample throughput. The instrument is affordable with friendly workflow and automation. Thus, the MiSeq platform is suitable for the sequencing of small genomes, such as bacteria genomes from 0.3 to 15 Gb using paired libraries. The read length is up to 300 bp. Meanwhile, the NextSeq, HiSeq, and NovaSeq instruments with considerable prices are applied for much higher throughput. Of which the HiSeq 2500 platform is the best option for high-throughput sequencing, as the high capacity of generating reaches more than 1 Tb sequence with a read length of 250 bp using paired libraries. In addition, other two versions, HiSeq 3000 and 4000 exhibit an approximate throughput of 125,750 Gb and 1,251,500 Gb, respectively, with the same maximum read length (150 bp) using paired libraries (Van Dijk et al. 2014). In 2014, HiSeq X Ten was released for population-scale whole genome sequencing (WGS) (Reuter et al. 2015). The system consists of new patterned flow cell technology and increases the throughput.

The Illumina sequencing approach generates a very large number of short sequence reads. For all Illumina models, the overall error rates are less than 1% (Reuter et al. 2015). However, in order to generate a consensus and to ensure high confidence in determination of the sequence differences, deep sequencing with more than ten-fold even coverage is required. Similar to the

conventional methods, deep sampling allows the use of weighted 'majority voting' and statistical analysis to identify homozygotes and heterozygotes and to distinguish sequencing errors.

12.2.4 AB SOLiD Sequencing Platform

The SOLiD (Supported Oligonucleotide Ligation and Detection) System 2.0 platform was developed in 2005. The Applied Biosystems SOLiD (AB SOLiD) sequencing technology was refined by Applied Biosystems (now named Life Technologies), and the instrument was released in 2007 (Voelkerding et al. 2009). SOLiD sequencing is a massive parallel and ligation-mediated sequencing method, based on the Polonator technology (Shendure et al. 2005). Similar to the 454 technology, DNA fragments are ligated to oligonucleotide adapters, attached to beads, and amplified using the emPCR in AB SOLiD platform (Shendure and Ji 2008; Voelkerding et al. 2009).

For template preparation, DNA molecules are sheared into fragments ranging from 400–850 bp in size. The ends of DNA fragments are repaired and ligated with 'P1' and 'P2' DNA adapters (Valouev et al. 2008). Emulsion PCR is performed to immobilize the sequencing library DNA. The amplicons are captured onto the 'P1' coated 1 μM paramagnetic beads. The immobilizing the modified beads generate the high-density and semi-ordered arrays to a glass flow-cell surface. The glass slides can be segmented up to eight chambers, in order to facilitate up the scaling of the analyzed samples.

In the in vitro sequencing by synthesis, a DNA ligase is applied. Each cycle applies partially degenerated and fluorescently-labeled DNA octamers with dinucleotide complement sequence recognition core. These detection oligonucleotides are hybridized to the template. Perfectly annealing sequences are ligated to the primer. After imaging, the labeled portions of the octamers are cleaved, while the unextended strands are captured (Valouev et al. 2008). A new cycle begins from the priming site (Shendure and Ji 2008).

The SOLiD system achieves a slightly better performance in terms of sequencing accuracy, due to the redundant sequencing of each base twice by a dinucleotide detection core structure of the octamer sequencing oligonucleotides. Tens of millions of 25–50 nt reads are generated by a single run. However, in the SOLiD system, the ligation-based method requires complex panel of labeled oligonucleotides, and sequencing proceeds by off-set steps. The interpretation of the raw data requires a complicated algorithm (Valouev et al. 2008).

Besides the above technologies, other NGS technologies, such as Ion Torrent PGM (acquired by Life Technologies, later by Thermo Fisher Scientific, USA), Polonator system (Dover/Harvard, USA) and single-molecule sequencer (Helicos BioSciences, USA) also have been developed in the past ten years (Harris et al. 2008; Quailet al. 2012). All the high-throughput DNA sequencing technologies involve sequencing library preparation, DNA immobilization onto a solid support, cyclic sequencing using automated fluidics devices, and detection of molecular events by imaging.

12.2.5 SMRT Sequencing

As to the read length in genome sequencing, most NGS technologies generate only a few hundred bases for the gradual intermolecular dephasing among DNA clones. Single-molecule sequencing technologies, i.e., the 3rd generation sequencing accomplish the long reads sequencing (Wang et al. 2015). Single-molecule real-time sequencing (SMRT) technology is a proprietary approach, which was commercialized by Pacific Biosciences (Pacific Biosciences, USA). Pacific Biosciences real-time sequencing (PacBio PSII) machine was released in 2010. RSII involves single-molecule real-time sequencing technology using fluorescent dye labeled nucleotides. Long DNA fragments of 20 kb and longer can be obtained in a few hours. Thus, the

SMRT platform is regarded as the most widely used third-generation sequencing technology (Heather and Chain 2016). Compared with the NGS technologies, SMRT sequencing exhibits many benefits on genome sequencing, such as generating longer average read lengths, higher consensus accuracy, uniform coverage, simultaneous epigenetic characterization, and single-molecule resolution. The SMRT platform involves preparing sequencing library, real-time sequencing, and imaging the continuous incorporation of dye-labelled nucleotides (Korlach et al. 2008; Travers et al. 2010).

In SMRT platform, the genomic DNA is randomly fragmented and then end-repaired. Single DNA oligonucleotide is used as adapter to form an intramolecular hairpin loop on either end of double-stranded DNA (dsDNA) region. The prepared DNA template is named SMRTbell template as sequencing libraries. The template molecule can be sequenced multiple times using a strand displacing polymerase. As a result, the clonal amplification is avoided (Reuter et al. 2015).

For SMRT sequencing, each SMRT chip are made of a 100 nm-thick metal film containing thousands of zero-mode waveguide (ZMW) cavities of 10–50 nm in diameter (Korlach et al. 2008) (Fig. 12.3). Within each ZMW detector, single DNA polymerase molecule is attached to the bottom surface. As the sequencing reaction begins, the tethered polymerase incorporates nucleotides with individually phospholinked fluorophores. Four fluorophores correspond to the four specific bases of the growing DNA chain (Korlach et al. 2008). When phospholinked nucleotides are incorporated into the single-stranded DNA molecule template, the single-molecule DNA polymerase activity is visualized using biotinylated fluorescent latex beads. As PacBio sequencing takes place in real time, kinetic variation interpreted from the light-pulse movie can be applied for the detection of methylated cytosines (Flusberg et al. 2010). Using the latest chemistry, each SMRT cell generates about 50,000 reads and up to 1 Gb data (Reuter et al. 2015).

To date, PacBio platform is more suitable for sequencing core facilities and to construct the scaffold of genomes because of the low run throughput, high sequencing cost per sample, expensive instrument, and large footprint (Table 12.1). PacBio SMRT has a higher inherent error profile (11–15%) than the short-read technologies. Even though the error rate can be reduced by increasing the number of sub-reads generated, the read length becomes shorter. In recent years, Pacific Biosciences released a higher throughput and more affordable version of RSII called Sequel.

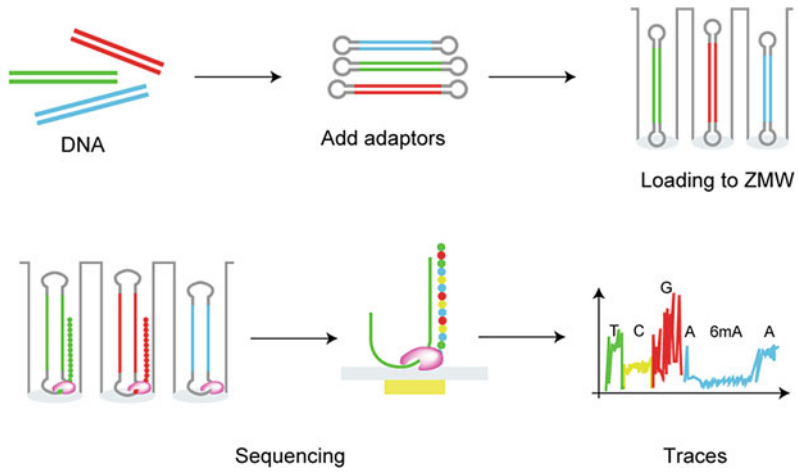


Fig. 12.3 PacBio genome analyzer sequencing

Table 12.1 Comparison of the main sequencing technologies used for genome sequencing

Platform or instrument	Throughput range (Gb)*	Read length (bp)	Strength	Weakness
1. *Sanger sequencing*				
ABI 3500/3730	0.0003	Up to 1 kb	Read accuracy and length	Cost and throughput
2. *Illumina*				
MiniSeq	1.7–7.5	1 × 75 to × 150	Low initial investment	Run and read length
MiSeq	0.3–15	1 × 36 to 2 × 300	Read length, scalability	Run length
NextSeq	10–120	1 × 75 to 2 × 150	Throughput	Run and read length
HiSeq (2500)	10–1000	× 50 to × 250	Read accuracy, throughput,	High initial investment, run
NovaSeq 5000/6000	2000–6000	2 × 50 to × 150	Read accuracy, throughput	High initial investment, run
3. *Ion Torrent*				
PGM	0.08–2	Up to 400	Read length, speed	Throughput, homopolymers
S5	0.6–15	Up to 400	Read length, speed,	Homopolymers
Proton	10–15	Up to 200	Speed, throughput	Homopolymers
4. *Pacific biosciences*				
PacBio RSII	0.5–1**	Up to 60 kb	Read length, speed (Average 10 kb, N50 = 20 kb)	High error rate and initial
Sequel	5–10***	Up to 60 kb	Read length, speed (Average 10 kb, N50 = 20 kb)	High error rate
5. *Oxford nanopore*				
MInION	0.1–1	Up to 100 kb	Read length, portability	High error rate, run length

*refers to the throughput ranges which are determined by available kits and run modes on a run base
**refers to throughput number per single-molecule real-time cell
***refers to the error rate (%) of reads which in turn results in false-positive variant calling

12.2.6 Oxford Nanopore Sequencing

Nanopore-based sequencing is another promising sequencing technology using single-molecule strategy. Nanopore analysis involves driving single molecule through a nanoscale pore, monitoring the ionic current through the nanopore changes, and identifying the bases (Venkatesan and Bashir 2011). The first nanopore sequencing device MinION was commercialized in 2014.

In nanopore platform, library preparation consists of DNA fragmentation and adapter ligation. DNA is sheared and end-repaired using NEBNext End Repair Module (Ashton et al. 2015). Ideally, the template molecule carries a motor enzyme, tether, hairpin, and hairpin motor (Ashton et al. 2015). Similar to SMRT sequencing, library preparation for nanopore sequencing can be performed without PCR amplification.

A sequencing flowcell comprises hundreds of independent micro-wells, each containing a synthetic bilayer perforated by biologic nanopores. Template strands are slowed down through the nanopore by a motor enzyme, while

the complement strands are slowed down by HP motor (Quick et al. 2014). Thus, the sequencing accuracy increases. A flowcell can obtain gigabases of sequence data (Table 12.1). Recently, a wild tomato species (*Solanum pennellii*) was sequenced using Nanopore sequencing. The average length of the genome reads ranged from 6 to 15 kb (Schmidt et al. 2017). Even though the ratio of the base insertion, deletion, and substitution in reads are high, nanopore sequencing is still believed as a promising sequencing technology for the huge data, long length, and high sequencing speed (Jain et al. 2015; Reuter et al. 2015).

12.2.7 Hi-C Sequencing

Chromatin conformation capture techniques provide a new method to detect the genome structure. Hi-C (high-throughput chromosome conformation capture) sequencing technology involves sequencing pairs of interacting DNA fragments and mapping the chromatin interactions of genome or genome regions (Servant et al. 2015). Differing from other sequencing technologies, the Hi-C technique could demonstrate the conservation of megabase-long domains and the probable physical proximity between pairs of chromosomal loci on a genome (Yaffe and Tanay 2011). During constructing the Hi-C sequencing library, chromatin is cross-linked with formaldehyde, and the genomic DNA is digested and fragmented using restriction enzymes. The interacting DNA fragments are re-ligated and form chimeric DNAs (Lieberman-Aiden et al. 2009; Belton et al. 2012). A biotin-labeled nucleotide is designed to incorporate in the chimeric DNA for purification. After are sequenced, the chimeric DNAs are mapped back to the genome (Belton et al. 2012). At present, Hi-C technique becomes a necessary process for genome assembly to chromosome scale.

12.2.8 Bionano Sequencing

The contiguity of the scaffolds affects the final quality and the accuracy of the genome assemblies. To obtain long reads, the BioNano Irys® System was innovated by BioNano Genomics Company (BioNano, the USA) (Daset al. 2010; Shelton et al. 2015). The updated Bionano Saphyr™ system was constructed in 2015. In Bionano sequencing system, the long DNAs (up to above 300 kb) are extracted and fluorescently labeled for forming single molecular physical maps. For a Bionano Saphyr chip, about 120,000 parallel nano channels array in silicon for linearizing long DNA molecules in solution. After sequenced, the obtained high throughput physical maps exhibit the order and location of long repeats, specific restriction enzyme sites, and long DNA molecules and ensure the concise assembly of the super scaffolds (Shelton et al. 2015) (https://www.prnewswire.com/news-releases/bionanos-irys-system-adopted-by-leading-genomics-centers-for-comprehensive-detection-of-human-genome-structural-variation-839245728.html). In order to create more contiguous consensus genome maps using both sequencing and BioNano map data, the Hybrid Scaffold tool was developed by BioNano Genomics company (Shelton et al. 2015). BioNano sequencing supplies a high- efficient and low- cost method for generating whole genome maps.

12.3 Hybrid Sequencing Strategy

Complicated sequencing approaches impel the genomics research in organism (Topol 2012). The large amount of sequence data output also enhances the development of bioinformatics (Voelkerding et al. 2009). As more and more genomes are sequenced, the flexible sequencing strategies become essential especially for large-

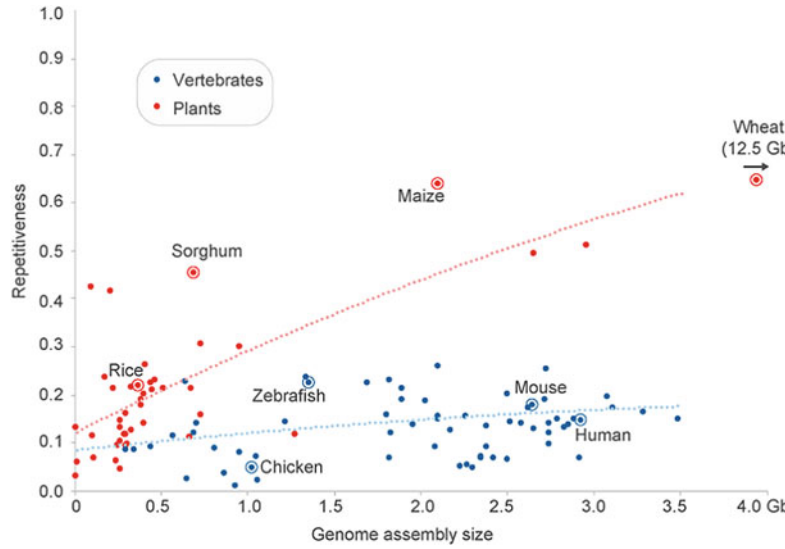


Fig. 12.4 Comparison of size and repetitiveness of plant and vertebrate genomes (Cited from Jiao and Schneeberger (2017))

genome-sized organisms. For plants and vertebrates, the genome size corresponds to the repetitiveness of genomes (Fig. 12.4). Up to the date of April, 30th, 2018, about 283 genomes of 273 plant species (including algae) have been sequenced and published (http://www.plabipd.de/timeline_view.ep). However, the large genome assemblies, such as maize (2300 Mb), barley (5100 Mb) and wheat (hexaploid, 17,000 Mb) are highly fragmented because of the inability of sequencing technologies to span complex repeat regions. Moreover, the ploidy, heterozygosity, and sequence coverage also affect the quality of genome assemblies in plant. To overcome the inherent errors of each data type and to increase the completeness of the genome sequences, hybrid sequencing strategy is recommended (Brown et al. 2014; Utturkar et al. 2014; Zimin et al. 2017).

According to the above introductions, each sequencing technology has its own strengths and weaknesses especially for large genome sequencing. For example, second-generation sequencing technologies generate the accurate and inexpensive short reads, but the assemblies are always fragmented and incomplete. Recent advances in single molecule real-time (SMRT) and nanopore sequencing technologies enables high-quality assemblies using long and inaccurate reads. However, these approaches require high coverage with high cost. Meanwhile, applying the 3rd-generation sequencing alone leads to the introduction of inherent errors in sequences for the relatively low accuracy. Therefore, combining the various sequencing technologies and making the perfect sequencing strategies for a specific species is crucial to facilitate the genomic research.

The hybrid sequencing approach involves integrity of the short and accurate NGS sequencing data with the long and inaccurate reads of the 3rd data. In order to resolve the complicated and high-repeated DNA segments, Utturkar et al. (2014) combined Illumina sequence data with PacBio data using various hybrid assembly strategies to assess the genome assembly quality of 4 bacterial species. The results proved that ALLPATHS-LG algorithm is ideal, as the assembled genome sizes and N50 values of the 4 bacteria was improved. On the other hand, the genome sequencing strategies developed from bacterial artificial chromosome (BAC) by BAC sequencing, whole genome shotgun (WGS), to chromosome sorting accompanied with BAC by BAC sequencing or whole genome shotgun sequencing also show the advantages of hybrid sequencing approaches (Bogler et al. 2014).

12.4 Sequencing Strategy of *S. Indicum* Genome

12.4.1 Estimation of the Genome Size of *S. Indicum*

The sequencing strategy of a genome is designed according to genome size, repeat content, heterozygous rate, and ploidy characters. Before starting the SGP, accurate estimation of the genome size and the chromosome set characteristics has been performed (Zhang et al. 2012, 2013). We firstly assessed the genome size of *S. indicum* using flow cytometry (BD FACSation ™, USA) (Doležel et al. 1989) (Fig. 12.5 and Table 12.2). The soybean variety William 82 was applied as genome size reference (Schmutz et al. 2010). The nuclear DNA content of *S. indicum* (var. Yuzhi 11) was calculated according to the formula: sample 2C value = reference 2C peak × sample 2C peak)/(reference 2C peak) 0.1 pg = approximately 965 Mb. The genome size of var. William 82 is 950 Mb, and the 2C DNA content is 0.98 pg.

The results showed that the average signal peak value of var. Yuzhi 11 was 20.44, lower than that of soybean var. Williams 82 (52.68). The average 2C DNA content of var. Yuzhi 11 was 0.38 pg/2C DNA, and the estimated genome size of *S. indicum* was 369 Mb (Zhang et al. 2013).

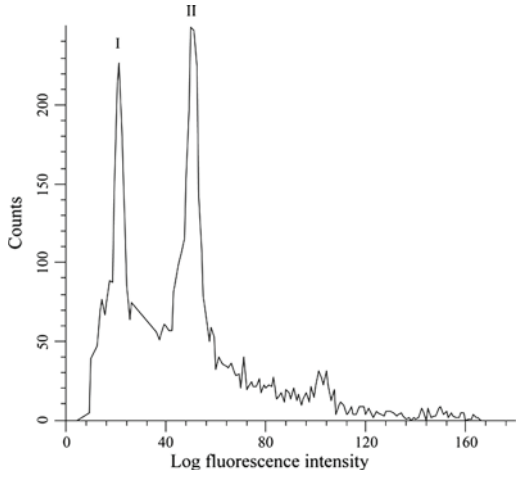


Fig. 12.5 2C DNA content estimation of *S. indicum*, var. Yuzhi 11 using flow cytometry. **a** 2C DNA peak signal of *S. indicum*, var. Yuzhi 11. **b** 2C DNA peak signal of soybean, cv. Williams 82. (Supplied by Haiyang Zhang)

On the other hand, we accurately estimated the genome size of *S. indicum* var. Yuzhi 11 by the K-mer frequency using 16.77 Gb of trimmed Illumina PE (500 bp) reads (Zhang et al. 2013). K-mer refers to the substring of length k. The occurrence time of k-mers indicates the sequencing coverage and reflects the genome size (Marçais and Kingsford 2011). In sesame, the distribution of K-mer follows a Poisson's distribution (Fig. 12.6). The peak k-mer frequency is 39, and the minimum point is 10.

In the study, the k-mer depth distribution of Illumina reads was calculated using Jellyfish (v1.1.4) (Marçais and Kingsford 2011) (Table 12.2). The results showed that the k-mer size of *S. indicum* is 17, and the estimated genome size is 354 Mb. The above two groups of estimation results using flow cytometry and k-mer depth distribution, respectively are similar. The genome size (354–369 Mb) of sesame is relatively smaller than the genome (430 Mb) of *Oryza sativa* L. and reflects the relative uncomplicated character.

12.4.2 Sequencing Strategy

Implement of the Sesame Genome Project spans the specific blooming stage of genome sequencing technologies. As many new commercial sequencing technologies are gradually developed and applied, the sequencing strategy of *S. indicum* is updated accordingly. At the initiation stage of the SGP (in 2010–2011), the SGWG referenced the published genomes of papaya (Ming et al. 2008), cucumber (Huang et al. 2009), and Potato (Xun et al. 2011) and designed the sequencing strategy, 'ABI3730xL +Roche/454 + Illumina/Solexa' (i.e., version 1). This genome sequencing strategy for sesame involves the contemporary advanced sequencing technologies of Solexa pair-end sequencing and Roche-454 pyro sequencing platforms, as well as conventional Sanger sequencing technologies at that time (Zhang et al. 2013). There are four steps

Table 12.2 Genome size estimation of *S. indicum* using flow cytometry

Sample no	Sample 2C peak value[a]	reference 2C peak value[b]	CV (%)	2C DNA content (pg)	Genome size (Mb)
1	21.72	53.60	3.94	0.40	385
2	20.63	51.31	4.15	0.40	386
3	19.72	52.28	2.15	0.37	358
4	19.86	51.83	4.59	0.38	364
5	20.53	54.61	2.90	0.37	358
6	20.17	52.44	3.75	0.38	365
Mean ± SD	20.44 ± 0.72	52.68 ± 1.22	3.58 ± 0.9	0.38 ± 0.01	369 ± 11.69

[a]Sesame cultivar Yuzhi 11
[b]Soybean var. William 82 as the reference. The genome size is 950 Mb and the 2C DNA content is 0.98 pg (Supplied by Haiyang Zhang)

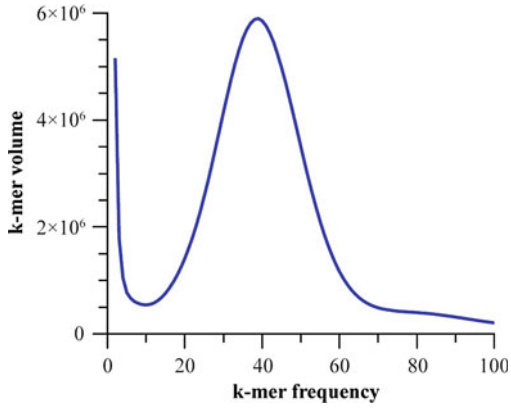


Fig. 12.6 Kmer distribution of the genome of *S. indicum* var. Yuzhi 11. The 16.77 Gb Illumina data with the genome coverage of 98× is used for analysis (Cited from Zhang et al. (2013))

to perform the sesame genome, i.e., (1) applying Solex and Roche 454 sequencing platforms to generate short length reads; (2) sequencing both ends of BAC library using ABI 3730 to increase the genome coverage and to fill the gaps; (3) performing the BAC pooling strategy using NGS sequencing technologies to generate the whole sequences of thousands of BAC clones in gap regions and to form super scaffolds; and (4) applying the whole genome walking with BAC libraries using 3730 platform to increase the genome coverage and to avoid of the possible assembly errors of the homologous chromosomes. The main sequencing objectives are: (1) the assembled genome covers 90% of the whole genome; (2) the total number of scaffolds is less than 500; (3) the N50 of the scaffolds is more than 5 Mb. Moreover, the molecular genetic map and optimal map would be applied for sesame genome assembly.

In 2012, the first draft genome of sesame (var. Yuzhi 11) was assembled using 98 Gb (276×) Solex data alone (Zhang et al. 2013). The gnome assembly length was 293.7 Mb, occupying 82.9% of the estimated genome size. The N50 of scaffolds was 22.6 kb. The N90 of scaffolds was 4.3 kb. More analysis using 'ABI3730xL + Roche/454 + Illumina/Solexa' sequencing data reflected that the N50 of scaffold in sesame reached 1.7 Mb, which was longer than 1.3 Mb of the published potato genome in 2011 (unpublished data, Haiyang Zhang).

In order to increase the quality of the sesame genome assembly, PacBio SMRT sequencing technologies were introduced into the SGP. Genome sequencing and assembly strategies of SGP were updated accordingly. In the new sequencing strategy (version 2), 96 SMRT cells were sequenced on PacBio RSII platform to generate long reads. Meanwhile, the sequencing task of the BAC clones using NGS sequencing technologies was terminated to decrease the sequencing expense (partially shown in Fig. 12.7). These 3rd-generation data are used for scaffold assembly. Subsequent assembly results proved that the new sequencing strategy

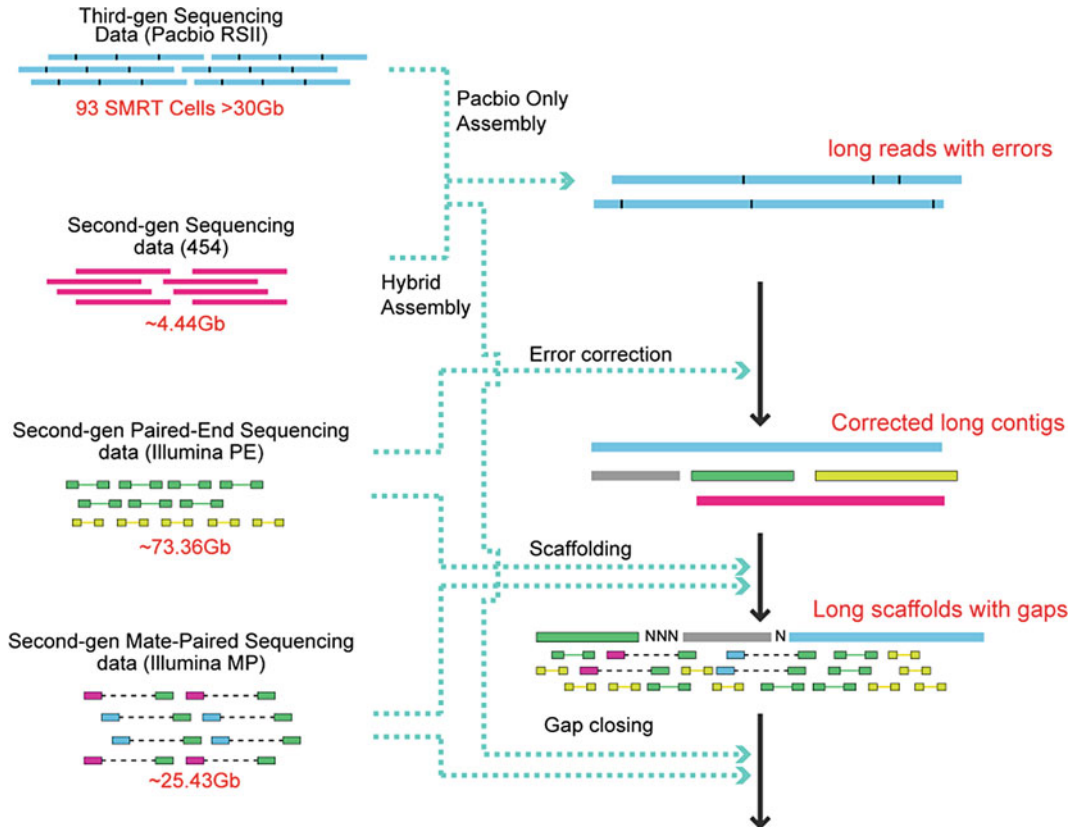


Fig. 12.7 Updated genome sequencing strategy (version 2) of *S. indicum* (Provided by Haiyang Zhang)

was helpful to assemble a fine sesame genome map with high completeness and accuracy (Miao 2014; Miao and Zhang 2016).

Furthermore, in order to assemble a chromosome-scaled genome map for sesame, optimal mapping and Hi-C sequencing were tried. As a result, Hi-C sequencing was successfully applied for the genome assembly (please see in the Chap. 13). The assembly results reflect the importance of the proper genome sequencing and assembly strategy.

References

Anderson S (1981) Shotgun DNA sequencing using cloned DNase I-generated fragments. Nucl Acids Res 9(13):3015–3027

Ashton PM, Nair S, Dallman T, Rubino S, Rabsch W et al (2015) MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. Nat Biotechnol 33(3):296

Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y et al (2012) Hi-C: a comprehensive technique to capture the conformation of genomes. Methods 58:268–276

Berlin K, Koren S, Chin C, Dreke JP, Landolin JM et al (2015) Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. Nat Biotechnol 33:623–630

Bolger ME, Weisshaar B, Scholz U, Stein N, Bjorn U et al (2014) Plant genome sequencing-applications for crop improvement. Curr Opin Biotechnol 26:31–37

Brown SD, Nagaraju S, Utturkar S, Tissera SD, Segovia S et al (2014) Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of *Clostridium autoethanogenum* and analysis of CRISPR systems in industrial relevant Clostridia. Biotechnol Biofuels 7(1):40

Cheng S, Melkonian M, Smith SA, Brockington S, Archibald JM et al (2018) 10KP: a phylodiverse genome sequencing plan. GigaScience 7(3):giy013

Das S, Vikato H, Hassibi A (2010) Model-based sequential base calling for Illumina sequencing. In: IEEE international workshop on genomic signal processing

and statistics (GENSIPS). Cold Spring Harbor, NY, pp 1–4
Das K, Tan P (2013) Molecular cytogenetics: recent developments and applications in cancer. Clin Genet 84(4):315–325
Doležel J, Binarová P, Lcretti S (1989) Analysis of nuclear DNA content in plant cells by flow cytometry. Biol Plant 31(2):113–120
Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC et al (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Meth 7:461–465
Harris TD, Buzby PR, Babcock H, Beer E, Bowers J et al (2008) Single-molecule DNA sequencing of a viral genome. Science 320(5872):106–109
Huang S, Li R, Zhang Z, Li L, Gu X et al (2009) The genome of the cucumber, *Cucumis sativus* L. Nat Genet 41(12):1275
Heather JM, Chain B (2016) The sequence of sequencers: the history of sequencing DNA. Genomics 107(1):1–8
Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol 8(7):R143
Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B et al (2015) Improved data analysis for the MinION nanopore sequencer. Nat Meth 12(4):351
Jiao WB, Schneeberger K (2017) The impact of third generation genomic technologies on plant genome assembly. Curr Opin Plant Biol 36:64–70
Korlach J, Marks PJ, Cicero RL, Gray JJ, Murphy DL et al (2008) Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. Proc Natl Acad Sci USA 105(4):1176–1181
Leamon JH, Lee WL, Tartaro KR, Lanza JR, Sarkis GL et al (2003) A massive parallel PicoTiterPlate based platform for discrete picoliter-scale polymerase chain reactions. Electrophoresis 24:3769–3777
Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289–293
Loman NJ, Quick J, Simpson JT (2015) A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Meth 12(8):733–735
Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27(6):764–770
Margulies M, Egholm M, Altman WE, Attiya S, Bader JS et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437(7057):376–380
Metzker ML (2010) Sequencing technologies-the next generation. Nat Rev Genet 11(1):31–46
Miao H (2014) The sesame genome project and sesame genome sequencing. plant and animal genome XXII conference. In: Plant and animal genome, 10–15th January, San Diego
Miao H, Zhang H (2016) The Genome of *Sesamum indicum* L. In: Plant and animal genome XXIV conference, 9–13th January, San Diego
Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte et al (2008) The draft genome of the transgenic tropical fruit tree 4 (*Carica papay*a Linnaeus). Nature 452 (7190):991–996
Mitchelson KR (2007) New high throughput technologies for DNA sequencing and genomics. Science press, Beijing, China
Nyrén P, Lundin A (1985) Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. Analyt Biochem 151(2):504–509
Quail MA, Miriam S, Paul C, Otto TD, Harris SR et al (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific biosciences and Illumina MiSeq sequencers. BMC Genom 13(1):341
Quick J, Constantinidou C, Pallen MJ, Oppenheim B, Loman NJ et al (2014) Draft genome sequence of *Elizabethkingia meningoseptica* isolated from a traumatic wound. Genome Announce 2(3):e00355-e414
Reuter JA, Spacek DV, Snyder MP (2015) High-throughput sequencing technologies. Mol Cell 58 (4):586–597
Rhoads A, Au KF (2015) PacBio sequencing and its applications. Genom Proteom Bioinformat 13(5):278–289
Shendure J, Ji H (2008) Next-generation DNA sequencing. Nat Biotechnol 26(10):1135–1145
Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O et al (2011) The genome of woodland strawberry (*Fragaria vesca*). Nat Genet 43(2):109
Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR et al (1977) Nucleotide sequence of bacteriophage ｜[phi]｜ X174 DNA. Nature 265(5596):687–695
Schmidt MHW, Vogel A, Denton AK et al (2017) De novo assembly of a new *Solanum pennellii* accession using nanopore sequencing. Plant Cell 29(10):2336–2348
Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T et al (2010) Genome sequence of the paleopolyploid soybean. Nature 463(7278):178–183
Shelton JM, Coleman MC, Herndon N, Lu N, Lam ET et al (2015) Tools and pipelines for BioNano data: molecule assembly pipeline and FASTA super scaffolding tool. BMC Genom 16:734
Shendure J, Porreca GJ, Reppas NB, Lin X, Mccutcheon et al (2005) Massively parallel sequencing. Nature 437 (7057):326–327
Sulston J, Du Z, Thomas K, Wilson R, Hillier L et al (1992) The *C. elegans* genome sequencing project: a beginning. Nature 356(6364):37–41
Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C et al (2015) HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol 16 (1):259
Topol EJ (2012) Comment on "the predictive capacity of personal genome sequencing". Science Transl Med 4 (135):135le3

Travers KJ, Chen-Shan C, Rank DR, Eid JS, Turner SW (2010) A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res 38(15):e159

Tuskan GA, Difazio S, Jansson S, Bohimann J, Grigoriev I et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313 (5793):1596–1604

Utturkar SM, Klingeman DM, Land ML, Schadt CW, Doktycz MJ et al (2014) Evaluation and validation of de novo and hybrid assembly techniques to derive high-quality genome sequences. Bioinformatics 30 (19):2709–2716

Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S et al (2008) A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. Genome Res 18(7):1051–1063

VanBuren R, Bryant D, Edger PP, Tang H, Burgess D et al (2015) Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*. Nature 527(7579):508–511

Van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C (2014) Ten years of next-generation sequencing technology. Trends Genet 30(9):418–426

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ (2001) The sequence of the human genome. Science 291(5507):1304–1351

Venkatesan BM, Bashir R (2011) Nanopore sensors for nucleic acid analysis. Nat Nanotechnol 6(10):615–624

Voelkerding KV, Dames SA, Durtschi JD (2009) Next-generation sequencing: from basic research to diagnostics. Clin Chem 55(4):641–658

Wang G, Liu Y, Zhu D, Klau GW, Feng W (2015) Bioinformatics methods and bilogical interpretation for next-generation sequencing data. Biomed Res Intl Atical ID:690873

Xun X, Pan S, Cheng S, Bo Z, Visser RGF (2011) Genome sequence and analysis of the tuber crop potato. Nature 475(7355):189–195

Yaffe E, Tanay A (2011) Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. Nat Genet 43:1059–1065

Yura T, Mori H, Nagai H, Nagata T, Ishihama A et al (1992) Systematic sequencing of the Escherichia coli geneme: analysis of the 0–2.4 min region. Nucleic Acids Res 20(13):3305–3308

Zhang H, Miao H, Li C, Wei L, Ma Q (2012) Analysis of sesame karyotype and resemblance-near coefficient. Chin Bull Bot 47(6):602–614

Zhang H, Miao H, Wang L, Qu L, Liu H, Wang Q (2013) Genome sequencing of the important oilseed crop *Sesamum indicum* L. Genome Biol 14(1):401

Zimin AV, Puiu D, Luo MC, Zhu T, Koren S et al (2017) Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome Res 27(5):787–792

Sesame Genome Assembly

13

Hongmei Miao, Yamin Sun, Lei Wang, and Haiyang Zhang

Abstract

Genome assembly strategy is a crucial step for reconstructing high-quality genomes in any genome project. With the continuous innovation of new sequencing platforms, various methods of genome assembly have been developed and applied in genome assembly. In order to construct a fine genome map for sesame, the Sesame Genome Working Group (SGWG) group analyzed and optimized the assembly strategy for sesame according to the various sequencing data and genome characteristics. The comprehensive genome assembly strategy for sesame comprised the next generation sequencing (NGS), the third assembly platforms, and Hi-C assembly strategy with supplementation of high-dense single nucleotide polymorphism (SNP) genetic map and bacterial artificial chromosome—fluorescent in situ hybridization (BAC-FISH) cytogenetic map. Comparison results of the specific and high-efficient assembly techniques in sesame were discussed in this chapter. The final assembled genome of sesame (cv. Yuzhi 11) comprises of 335 megabase (Mb) sequences with high completeness. About 313 Mb sequences are assembled into the 13 chromosome molecules. High accuracy and integrity of the assembled genome indicates the suitable genome assembly for sesame. The results provide an extraordinary example of the application of high-quality genome assembly techniques in higher plants with large genomes.

13.1 Introduction

Genome assembly is the most important process for reconstructing a genome and deciphering the DNA sequence of a species using short sequencing reads data. Since the next-generation sequencing (NGS) technology has been innovated and applied in genome sequencing in 2007, how to handle the structure of high-throughput and short-read-length data and to choose an appropriate assembly algorithm become a challenge to bioinformatics (Mitchelson 2007; Li et al. 2010a). Different sequencing methods produce vast and different reads in length and type. Different assembly methods should be chosen for combining different data based on the size and complexity of genomes and the obtained reads data. For example, in de

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

Y. Sun · L. Wang
TEDA School of Biological Sciences and Biotechnology, Nankai University, Tianjin, P. R. China
e-mail: nksunyamin@aliyun.com

L. Wang
e-mail: wanglei@nankai.edu.cn

H. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-98098-0_13

novo genome assembly, ABySS (Simpson et al. 2009) and SOAPdenovo (Li et al. 2010b) have been believed as the two most popular assemblers using De Bruijn graph (DBG) approach. Considering the various sequencing platforms and high throughput genome sequencing data (see in the Chap. 12) applied for the Sesame Genome Project, four different assembly methods, i.e., SOAPdenovo (Li et al. 2010a), DBG2OLC (Ye et al. 2016), Falcon (https://github.com/PacificBiosciences/FALCON/), and SmartDenovo (https://github.com/ruanjue/smartdenovo) were compared and applied for sesame genome assembly. The above main assemblers and algorithm approaches applied in large genome assembly are discussed in this chapter.

13.2 De Novo Assembly Algorithm

The de novo genome assembly process refers to combining reads via base pairing and independently reconstructing the original genome from other reference sequences using computer software (Sohn and Nam 2018). The size, topological complexity, and nonrandomness of genome sequences individually cause challenges in de novo assembly. Different assembly algorithms, such as greedy algorithms, overlap layout consensus (OLC), and DBG (such as Hamiltonian and Eulerian) are developed to cope with the complication of a genome and massive amounts of data computation (Miller et al. 2010; Ekblom and Wolf 2014; Wojcieszek et al. 2014; Sohn and Nam 2018).

The greedy algorithm approach is the most efficient method and has been applied in SSAKE, SHARCGS, and VCAKE (Verified Consensus Assembly by K-mer Extension) (Dohm et al. 2007; Jeck et al. 2007; Warren et al. 2007).

For short length reads assembly, the DBG approach is the most popular approach and has been largely applied in so many different assemblers, such as ALLPATHS-LG, ABySS, SOAPdenovo, and EulerSR (Imelfort and Edwards 2009; Miller et al. 2010; Nagarajan and Pop 2013; Sohn and Nam 2018). The approach consists of the steps: chopping reads into k-mer size fragments representing read length; forming a DBG using all the short k-mers; finding overlaps of k-1-mers, and inferring the consensus sequences on the DBG (Li et al. 2012).

Meanwhile, the OLC is also a common approach and has been applied in Newbler, Mira, and Edena (Chevreux et al. 2004; Hernandez et al. 2008; Reinhardt et al. 2008). The OLC consists of three steps, i.e., finding overlaps in a set of reads; constructing a graph based on the overlap information; and inferring consensus sequences through an intuitionistic assembly algorithm (Li et al. 2012). To combine reads data produced by 454 Life Sciences (Roche) and Illumina sequencing platforms, Reinhardt et al. (2008) developed Newbler and applied to assemble the de novo genome for rice for the first time.

13.3 Genome Assembly Software

13.3.1 SOAPdenovo

The SOAPdenovo program generates high-quality genome sequences from billions of Solexa reads of length 75 bp or less length. SOAPdenovo2 (Luo et al. 2012) implements an algorithm based on the sparse k-mer and reduces the required memory by up to 35 Gb for assembly of a whole genome for human. As a result, the assembly quality is improved. For SOAP denovo assembler, each step of de novo assembly from error correction, assembly of contigs and scaffolds, to gap filling) can be separately performed and simplified (Sohn and Nam 2018).

13.3.2 Newbler

Newbler is popular software using Overlap layout consensus (OLC) approach and distributed with the 454 sequencing machines by 454 Life Sciences. The first released version targeted unpaired reads in approximately 100 bp generated by the GS 20 machine. The Celera Assembler is a Sanger-era OLC assembler revised for 454 data.

13.3.3 Velvet

Velvet is one of the Eulerian de Bruijn graph assembler (Zerbino and Birney 2008). The software targets de novo assembly from short reads with paired ends from the Solexa platform. An extension allows it to assemble data sets composed solely of SOLiD reads (www.solidsoftwaretools.com). Based on bidirectional de Bruijn graph, the potential assembly errors could be corrected by so-called 'tour-bus' algorithms in assembly graph level.

13.3.4 DBG2OLC

High error rate of long sequencing reads makes the assembly of large genomes to be a challenge. DBG2OLC is designed to assemble the third generation sequencing (3GS) data (i.e., PacBio SMRT sequencing data) with the NGS data (Ye et al. 2016). Differing from the Hierarchical Genome-Assembly Process (HGAP) 13 which only use PacBio SMRT sequencing data, DBG2OLC is a hybrid assembly approach and is widely used to assemble mammalian-sized genomes with low consuming cost. Till now, DBG2OLC is still a popular assembler (Ye et al. 2016).

13.3.5 FALCON

In recent years, the implementation of the various high throughput sequencing technologies has impacted on basic genomic research. The similar haplotypes is also a challenge to de novo genome assembly. FALCON-Phase is designed to reconstruct contig-length phase blocks using Hi-C reads. The Falcon assembler shares many features with PBcR, such as raw reads overlapping for base error correction using DALIGNER and overlap filtering. In the FALCON-Phase pipeline, long scaffolds can be extended, and chromosome-scaled blocks can be obtained (Kronenberg et al. 2018) (https://github.com/phasegenomics/FALCON-Phase). Thus, Falcon approach becomes popular for fine genome assembly.

13.4 Hybrid Approach Strategy

A large amount of sequence data output enhance the development of bioinformatics (Voelkerding et al. 2009). Repetitive sequences and errors in the sequencing data can result in incomplete or biased coverage of the genome and misassemblies (McGrath 2007). Next-generation sequencing (Illumina) technology produces the inexpensive and short reads and results in the accurate but short contigs. Third-generation sequencing technologies such as single molecule real-time (SMRT) and nanopore sequencing platforms produce long but inaccurate reads, and result in long sequence assembly with high error rate. In order to perform a high genome assembly using nanopore data, some assemblers using different algorithms and error correction methods have been designed (Table 13.1). Thus, to combine reads from various technologies into a high-quality genome assembly, multiple assembly approaches have been applied to overcome the challenges.

One hybrid approach to genome assembly involves short and accurate second-generation sequencing data (such as from IonTorrent, Illumina or Roche 454) with long and inaccurate third generation sequencing data (such as from PacBio RS and nanopore) to resolve complex repeated DNA segments. For example, Koren et al. (2012) developed the PacBio corrected Reads (PBcR) approach and applied to correct the errors in long reads from Bio RS instrument using short reads. The long-read correction approach achieved >99.9% base-call accuracy, and realized the better assembly than non-hybrid strategies. Goodwin et al. (2015) developed an open-source error correction algorithm Nanocorr, which was specifically applied for correcting the hybrid errors (ranging from about 5 and 40% error) of Oxford nanopore reads. The error correction method realized a hybrid error correction of the nanopore

Table 13.1 List of assemblers for Oxford Nanopore MinION long reads

Assembler name	Algorithms	Error correction	Website
PBcR	HGAP or BLASR and Celera OLC	PBcR	http://wgs-assembler.sourceforge.net/wiki/index.php/PBcR
Canu	MHAP and Celera OLC	Canu	https://github.com/marbl/canu
Falcon	String graph and Celera OLC	Falcon	https://github.com/PacificBiosciences/falcon
ALLPATHS-LG	Porde Bruijn graph	ALLPATHS-LG	https://www.broadinstitute.org/software/allpaths-lg/blog/?page_id=12
SPAdes	De Bruijn graph	SPAdes	http://bioinf.spbau.ru/spades

reads using MiSeq data and formed the accurate de novo assembly of *Saccharomyces cerevisiae*. The contig N50 length was more than ten times greater than an Illumina-only assembly with >99.88% consensus identity.

In addition, the combination assembly approach of using a physical map and bacterial artificial chromosome (BAC) sequencing, as well as with whole genome shortgun has also been employed for Rat Genome Sequencing Project (Borwn Norway rat) (Mullins and Mullins 2004) and the domestic chicken genome assembly (International Genome Sequencing Consorotium) (Gill et al. 2004).

13.5 Assembly Technologies for Sesame Genome

13.5.1 De Novo Assembly of Illumina Data

In human genome project, two different assembly approaches i.e., a whole-genome assembly and a regional chromosome assembly, were applied and assembled about 2.91 billion base pairs consensus sequences representing the information of the 23 pairs of chromosomes of the *Homo sapiens* genome (Venter et al. 2001). In the Sesame Genome Project, in order to assemble a high-quality sesame genome (var. Yuzhi 11), the Sesame genome Working Group (SGWG) group initially performed the de novo assembly of 98 Gb Illumina data using SOAP de novo (Li et al. 2010b), Velvet (Zerbino and Birney 2008), and ABySS (Simpson et al. 2009), respectively, based on the de Bruijn graph approach (Li et al. 2012). The k-mer was optimized accordingly. As the K-mer was 67, the results proved SOAP de novo is the best software for sesame genome assembly (unpublished data, Haiyang Zhang). The relationship of Illumina data amount and the assembly size was analyzed. It was found that the genome assembly quality was improved with the increase of genome coverage (Zhang et al. 2013).

13.5.2 Mixed Assembly of Illumina and Roche-454 Data

In order to establish the specific and efficient genome assembly techniques for sesame, we firstly perform the mixed assembly using Illumina (98 Gb) and Roche-454 data (4.5 Gb) (unpublished data, Haiyang Zhang). The DB graph approach was applied to assemble the Illumina data. Then the Illumina assembly fragments were transferred into overlapped long sequences. Finally, the Roche-454 sequences were assembled with the overlapped long sequences for genome assembly. The parameters of the workflow and Newbler software were optimized during the process. The results showed that the length of contig N50 and scaffold N50 significantly increased with the addition of Roche-454 data. The assembled genome size increased from 231.5 to 262.4 Mb (Table 13.2) (unpublished data, Haiyang Zhang).

Table 13.2 Comparison of hybrid genome assembly strategies using various sequencing data for sesame

Assembly statistics for *S. indicum*																						
Sequencing data																						
Illumina data											√	√	√	√	√	√	√	√	√	√	√	√
PacBio data																						
10× depth	√											√										
20× depth		√											√									
30× depth			√											√								
40× depth				√											√							
50× depth					√											√						
60× depth						√											√					
70× depth							√											√				
80× depth								√											√			
90× depth									√											√		
100× depth										√											√	
Roche454 data																						√
Genome assembly																						
Assembled genome size (Mb)	85.6	67.8	139.4	208.3	260.9	278.6	298.4	304.9	327.1	328.4	231.5	225.9	251.1	261.3	269.1	277.1	283.4	293.0	301.4	309.5	317.5	262.4
Length of Contig N50 (Kb)	6.3	12.7	10.2	30.6	182.8	240.4	472.9	490.1	672.4	691.6	16.7	62.3	248.4	294.7	292.2	266.5	266.0	220.2	198.5	175.9	163.1	67.8
Length of Scaffold N50 (Kb)	6.3	12.7	10.2	30.6	182.8	240.4	472.9	490.1	672.4	691.6	22.6	62.3	248.4	294.7	292.2	266.5	266.0	220.2	198.5	175.9	163.1	1,899.4
Longest scaffold (Kb)																						14,138
Number of scaffolds																						3729

√indicates the data chose for assembly
All the data in the column are supplied by Haiyang Zhang

13.5.3 Assembly Guidance of PacBio Data

As the PacBio sequencing technology was applied in the SGP, the SGWG group performed PacBio data assembly in 2014. However, the supplied assembly and error correction software such as PBcR and HGAP by the instrument company was mainly applied in micro-organism genome assembly. Considering the huge data volume, the SGWG applied the supercomputer 'Tianhe 1' (TH-1) (Tianjin, China) and overcame the incompatibility of software with the super-computer in 2015. Meanwhile, the blasr script for sesame genome data was developed and applied in the blasr parallel analysis. Consequently, comparison of PacBio and NGS assembly results using blasr was performed. The PacBio data was assembled and combined with the NGS assembly results for sesame using the super-computer calculation platform for the first time. Only 144 h is needed to fulfill the sesame genome assembly guidance using the PacBio data.

During optimizing the assembly guidance of PacBio data, the relationship of PacBio data amount and the assembly size was also analyzed in sesame (Table 13.2) (Unpublished data, Haiyang Zhang). It was found that the genome size and assembly quality could be improved, as the depth of PacBio sequencing increased from 10× to 100× . The 32 Gb PacBio data with the 90× depth was appropriate for high-quality sesame genome assembly.

During the genome assembly, with the aid of PacBio assembly results, contigs of NGS were extended according to the blasr. Moreover, the two scaffolds with connection relationship were linked and formed scaffolding. As shown in Table 13.2, the assembled genome size reaches to 317.5 Mb under the combination of PacBio assembly with Illumina data (100×). The length of contig N50 and scaffold N50 are equal and reach to 163.1 Kb.

13.6 Screening the De Novo Genome Assembly Strategies

As described in the last chapter about sesame genome sequencing technologies, the SGWG constructed the complicated genome sequencing strategy for sesame and finally obtained the first generation data (BAC ends Sanger sequencing of 40,266 BAC clones), NGS data (i.e., 98 Gb Illumina and 4.5 Gb Roche 454 sequencing data), the third generation data (32 Gb PacBio SMRT data), and optical mapping data (OpGen sequencing of 60 MapCard) via the various sequencing technologies.

In the initial stage, in order to optimize the genome assembly pipeline, the SGWG designed four combination groups of genome assembly for sesame: (1) Illumina and Roche 454; (2) Illumina added with Roche 454 and BAC ends Sanger; (3) Illumina added with Roche-454 and OpgGendata; and (4) Illumina added with Roche 454, BAC ends sequencing data, and OpgGen-data covering the first and the second generation sequencing data and optical mapping data (Table 13.3). The comparison results showed that optical mapping data gave high function in extending the length of the largest scaffold (up to 17.19 Mb in Combination 3) and decreasing the N50 number (low to 14 in Combination 3).

As PacBio SMRT data was obtained for sesame, the SGWG further designed three groups of genome assembly workflows, based on the optimized assembly technique and software. A bacterial artificial chromosome-fluorescent in situ hybridization (BAC-FISH) cytogenetic map with 210 BAC clones (described in Chap. 8) was constructed for sesame and applied in the workflows at the same time (Figs. 13.1, 13.2 and 13.3).

For Workflow 1 (Fig. 13.1), the technique problem of data parawise comparison of PacBio data and NGS data was solved. Illumina and Roche-454 data was hybrid assembled. BAC

Table 13.3 Comparison of hybrid assembly strategies for sesame genome using the first and the second generation sequencing data and optical mapping data

Item	Hybrid assembly groups for sesame genome			
	(1) Illumina + Roche 454	(2) Illumina + Roche 454 + BAC ends Sanger	(3) Illumina + Roche 454 + Optical mapping	(4) Illumina + Roche 454 + BAC ends Sanger + Optical mapping
Length of the largest scaffold (bp)	8,353,280	10,657,165	17,185,203	18,145,754
Number of large scaffolds	4633	3352	3653	3284
Bases in large scaffolds (bp)	261,976,889	265,235,955	270,134,534	272,299,749
Number of N50 scaffold	45	27	14	11
Length of N50 scaffold (bp)	1,709,299	2,341,520	7,395,225	9,273,033
Number of N90 scaffold	197	121	64	54
Length of N90 scaffold (bp)	218,088	357,031	417,437	409,132
GC content (%)	35.18	35.05	35.19	35.05

ends and PacBio data were applied in scaffolding, while optical mapping data was used to construct super scaffolds. At the last step, the BAC-FISH cytogenetic map was applied for chromosome anchoring of super scaffolds.

For Workflow 2 (Fig. 13.2), the technique problem of de novo assembly of PacBio data was solved using super computer 'TH- 1'. PacBio reads were corrected and applied to construct scaffolds using de novo assembly, while Illumina, Roche 454, and BAC ends data were used for scaffolding. Subsequently, the optical mapping data were added to construct super scaffolds, and the BAC-FISH cytogenetic map was used to anchor scaffolds to the chromosome set.

For Workflow 3 (Fig. 13.3), the technique problem of alignment of NGS assembly with PacBio de novo assembly was solved. De novo assembly of the PacBio sequencing reads was carried out using continuous long reads (CLR) following the Hierarchical Genome Assembly Process (HGAP) workflow (PacBioDevNet, Pacific Biosciences, the USA) for sesame. HGAP consists of three steps: preassembly, de novo assembly with Celera Assembler, and assembly polishing with Quiver. Before assembly using Celera assembler (CA) version 7.0 software, the PacBio Rs_PreAssembler.1 module was used to correct the errors in the raw data generated from PacBio RS platform, with default minimum

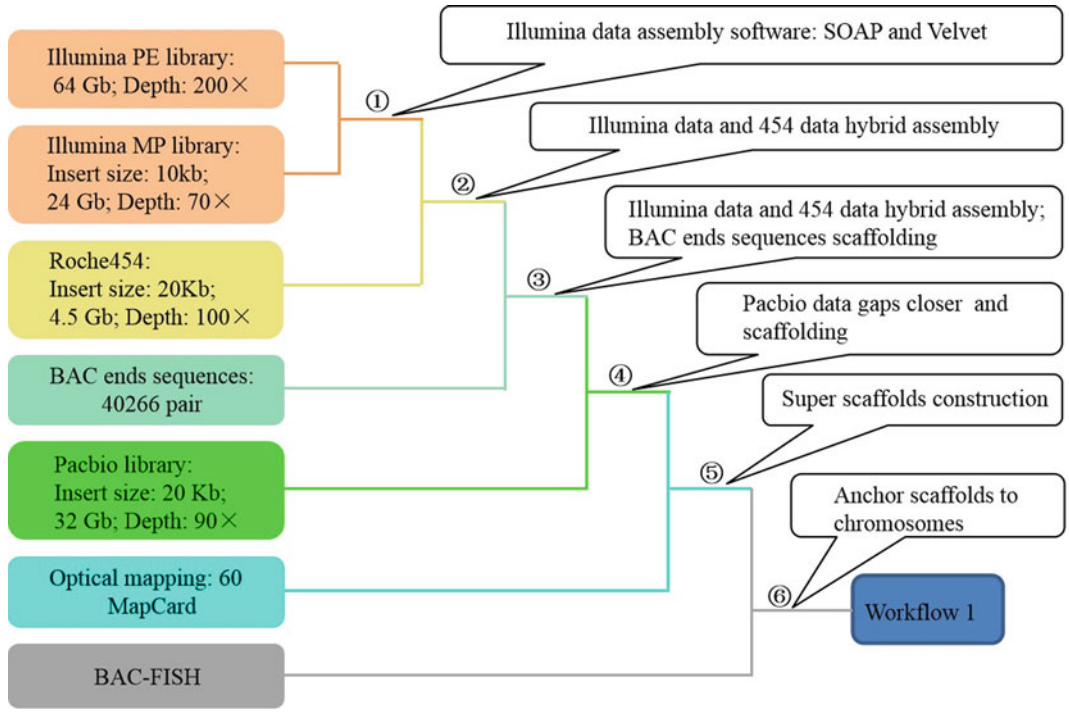


Fig. 13.1 Designed assembly workflow 1

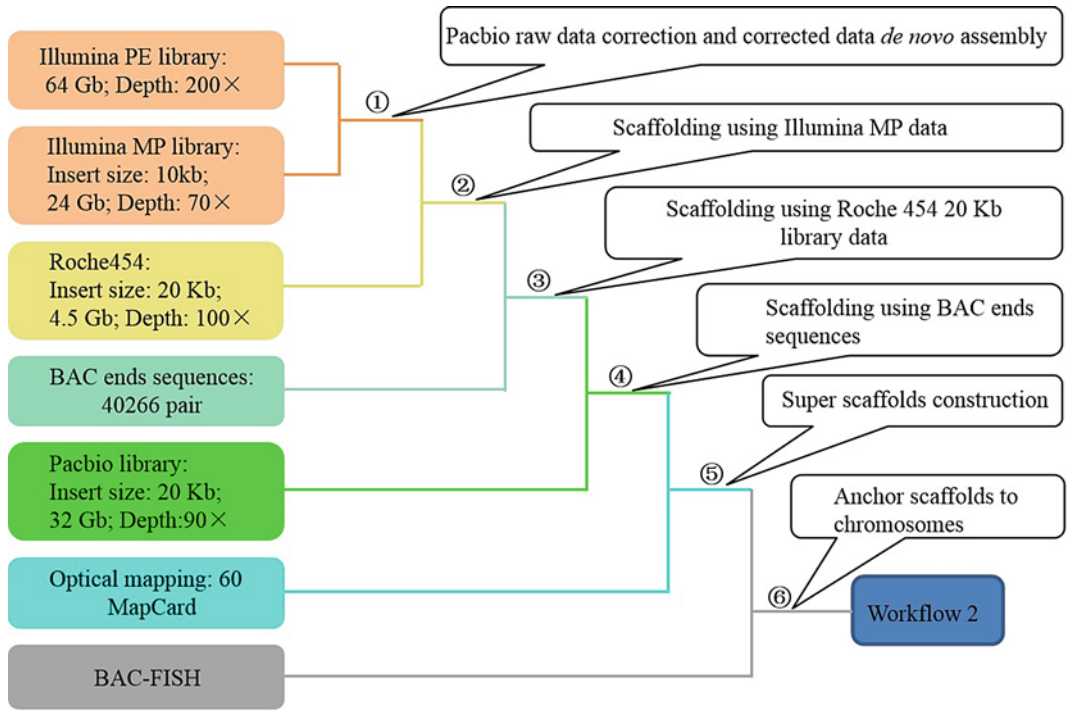


Fig. 13.2 Designed assembly workflow 2

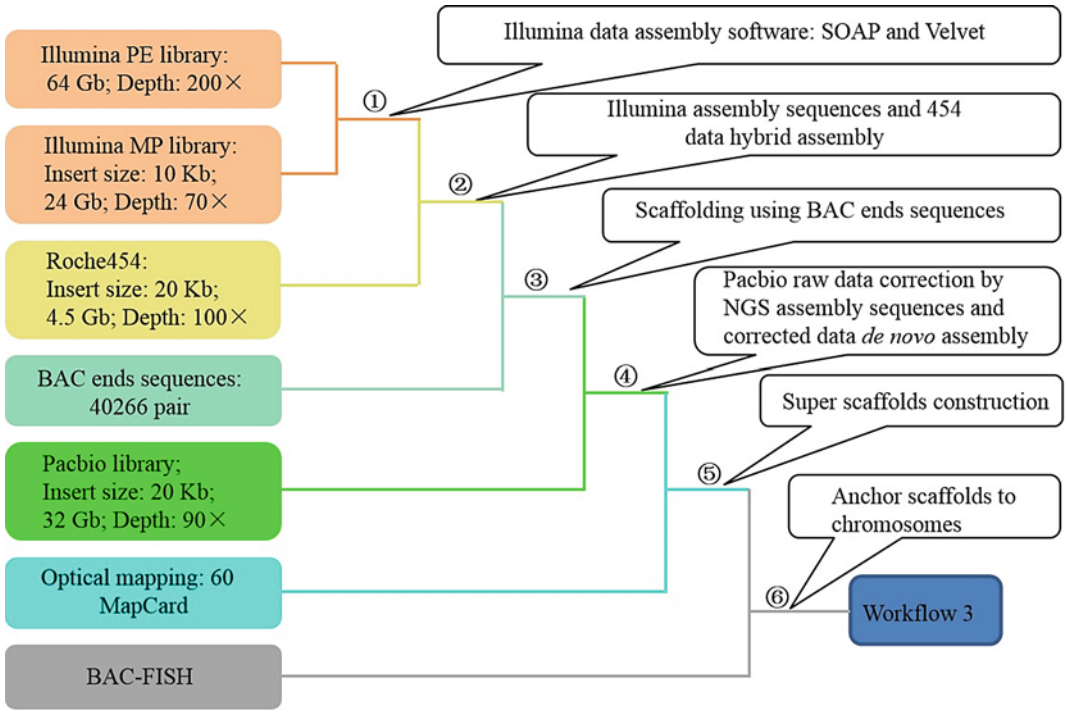


Fig. 13.3 Designed assembly workflow 3

subread length of 500 bp, a minimum read quality of 0.80, and a minimum subread length of 7500 bp. The read sequences of Mi-seq were mapped to polish assembly sequence from HGAP by BWA, and the single nucleotide polymorphisms (SNPs) and small Insertions and Deletions (InDels) were called and corrected by samtools and in-house script. Finally, scaffolds were generated, gap-filling was performed with meta-pair sequencing data (Illumina and Roche 454) using SSPACE 3.0 with default parameter values. The options and parameter values used in SSPACE were—×1-m 50-o 10-z 200-p 1.

Subsequently, the SGWG compared the assembly results of the hybrid assembly workflows with Illumina de novo assembly approach (Table 13.4). Differing from the results of SOAPdenovo for Illumina data assembly, other assemblers such as DBG2OLC, Falcon, and SmartDenovo using PacBio data or a hybrid of PacBio and Illumina data could obtain similar and satisfied assembly results. Comparison results indicated that Workflow 2 was the best strategy for sesame genome assembly. In Workflow 2, the large scaffolds and contigs reached to 319 Mb and 302 Mb, respectively. The largest scaffold was 13.88 Mb in length and presented the high-quality and long sequence assembly. The results reflected high efficiency of hybrid assembly approaches and the superiority of PacBio super long reads for large genome assembly. Especially, the genome assembly by Falcon displayed higher continuity than the other two methods.

Moreover, the SGWG also assembled the genome under the guidance of optimal mapping (OpgGen data), based on the new assembly of Workflow 2. The new assembled scaffolds

Table 13.4 Comparison of assembly results based on various assembly strategies

Assembly item	Illumina de novo	Workflow 1	Workflow 2	Workflow 3
Total scaffold number	2695	2525	3024	3303
Length of the largest scaffold (bp)	11,559,046	12,124,038	13,875,095	12,980,022
Bases in large scaffolds (bp)	275,386,645	278,345,234	319,899,578	272,752,635
Bases in large contigs	266,067,480	268,089,324	302,692,845	247,186,026
Number of N50 scaffolds	26	24	32	20
Length of N50 scaffold (bp)	2,866,350	3,067,089	3,942,549	4,126,754
Number of N90 scaffolds	123	120	156	81
Length of N90 scaffold (bp)	239,419	338,145	378,910	389,987
GC content (%)	35.108	35	34.235	35
N rate (%)	3.37	3.37	5.31	9.275

reached 321 Mb. The N50 length was 6.4 Mb and the N90 length of scaffolds was 418 kb (Unpublished data, Haiyang Zhang). Later, the SGWG performed Hi-C sequencing and added the Hi-C sequencing data into the sesame genome assembly in 2017 (data not shown) (Fig. 13.4). Considering the similar function of optimal mapping to the super dense SNP genetic map (Zhang et al. 2013) and BAC-FISH cytogenetic map (unpublished data, Haiyang Zhang) and the genome assembly results, the SGWG optimized and determined the final genome assembly for sesame: assembling 'ABI3730xL + Roche/454 + Illumina/Solexa + PacBio SMRT' sequencing data added with the 'SNP genetic map + BAC-FISH physical map + Hi-C library' assisted assembly in 2018 (Unpublished data, Haiyang Zhang). As a result, the first chromosome-scaled genome (version 3) for sesame (var. Yuzhi 11) was assembled (Table 13.5).

As to the final statistics for the sesame genome shown in Table 13.5, the assembled genome size is 335.19 Mb (containing artificial addition of N for gaps) with the largest scaffold of 29.5 Mb. The size above which 50% of the total length of the sequence assembly can be found (N50) of scaffolds is 7. The N50 length of scaffolds was 23.50 Mb (Unpublished data, Haiyang Zhang). Evaluation of protein-coding sequences and chromosome set comparison results showed that the assembled 13 genome molecules accord with the chromosome set of sesame (Zhao et al. 2018), reflecting the high quality and completeness (Unpublished data, Haiyang Zhang). Compared the de novo Illumina assembly (Zhang et al. 2013; Wang et al. 2014), the chromosome-scaled assembled genome supplies more precise information for genome structure and genome evaluation analysis and molecular genetics and genomics research in sesame.

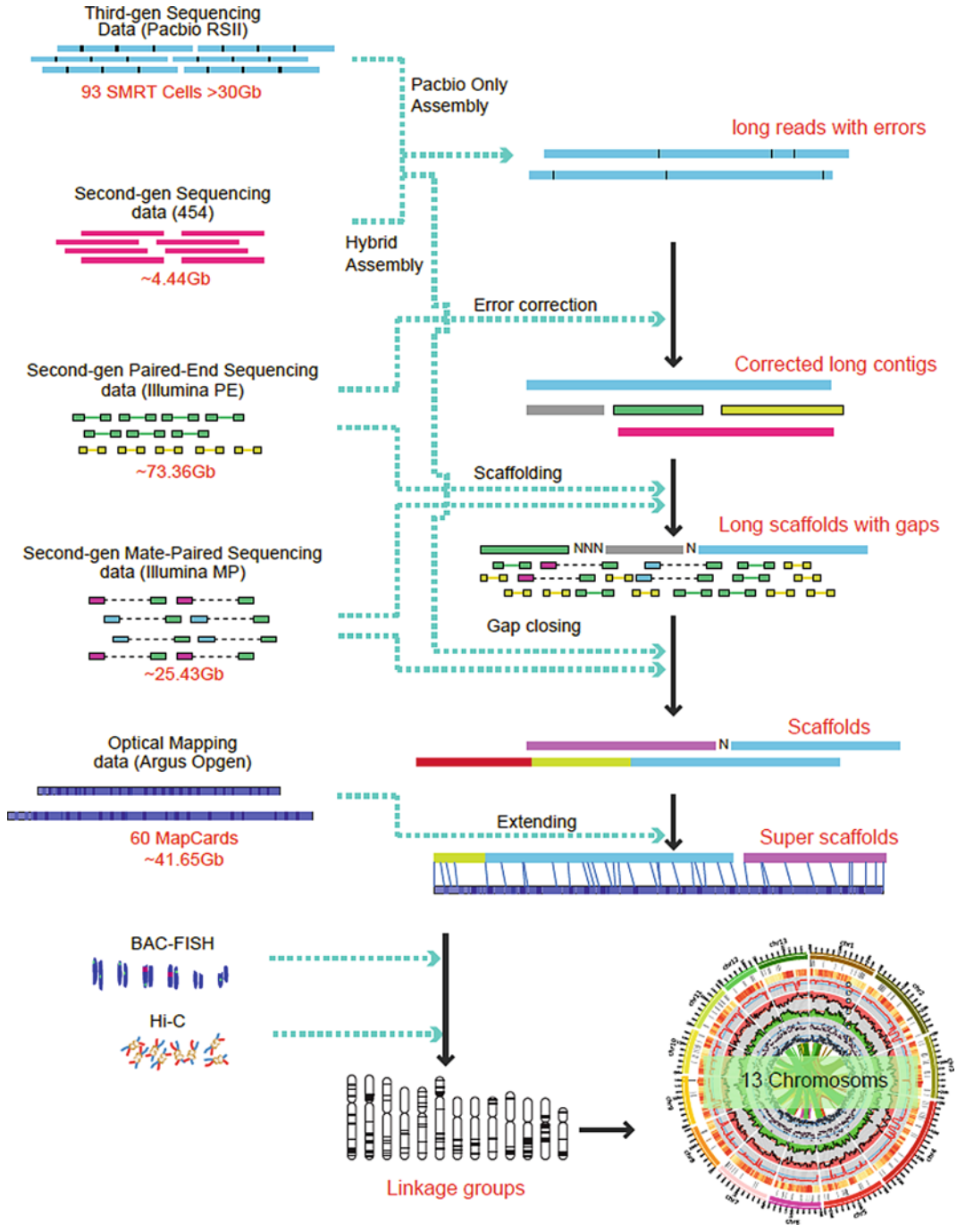


Fig. 13.4 Improved genome assembly workflow for sesame

Table 13.5 Information of the final genome assembly for sesame (var. Yuzhi 11)

Genome information		Sesame (var. Yuzhi 11)
Large scaffolds (>1000 bps)		
	Length of the largest scaffold (bp)	29,511,934
	Number of large scaffolds	2205
	Bases in large scaffolds (bp)	335,188,638
	Number of N50 scaffold	7
	Length of N50 scaffold (bp)	23,500,156
Total scaffolds		
	Scaffold number	2,210
	Total bases (bp)	335,184,649
Large contigs (>1000 bps)		
	Length of the largest contig (bp)	6,723,258
	Number of large contigs	3927
	Number of N50 contig	63
	Length of N50 contig (bp)	1,320,381
Total Contigs		
	Contig number	3927

References

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG et al (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res 14 (6):1147–1159

Dohm JC, Lottaz C, Borodina T, Himmelbauer H (2007) SHARCGS, a fast and highly accurate short-read assembly algorithm for de novo genomic sequencing. Genome Res 17(11):1697–1706

Ekblom R, Wolf JBW (2014) A field guide to whole-genome sequencing, assembly and annotation. Evol Appl 7(9):1026–1042

Gill BS, Appels R, Botha-Oberholster AM, Buell CR, Bennetzen JL et al (2004) A workshop report on wheat genome sequencing: international genome research on wheat consortium. Genetics 168 (2):1087–1096

Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Mccombie WR (2015) Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. Genome Res 25(11):1750–1756

Hernandez D, François P, Farinelli L, Osterås M, Schrenzel J (2008) De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. Genome Res 18(5):802–809

Imelfort M, Edwards D (2009) De novo sequencing of plant genomes using second-generation technologies. Brief Bioinformat 10(6):609–618

Jeck WR, Reinhardt JA, Baltrus DA, Hichenbotham MT, Jones CD (2007) Extending assembly of short DNA sequences to handle error. Bioinformatics 23 (21):2942–2944

Koren S, Schatz MC, Walenz BP, Martin J, Howard JT et al (2012) Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat Biotechnol 30(7):693–700

Kronenberg ZN, Hall RJ, Hiendleder S, Smith TPL, Kingan SB (2018) FALCON-phase: integrating PacBio and Hi-C data for phased diploid genomes. BioRxiv 327064

Li R, Zhu H, Ruan J, Qian W, Fang X (2010) De novo assembly of human genomes with massively parallel short read sequencing. Genome Res 20:265–272

Li Y, Hu Y, Bolund L, Wang J (2010) State of the art de novo assembly of human genomes from massively parallel sequencing data. Hum Genom 4(4):1–7

Li Z, Chen Y, Mu D, Yuan J, Shi Y (2012) Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and de-bruijn-graph. Brief Funct Genom 11(1):25–37

Luo R, Liu B, Xie Y, Li Z, Liu Y (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaSci 1(1):18

McGrath A (2007) Genome sequencing and assembly. Persp Bioanalys 2:327–355

Miller JR, Koren S, Sutton G (2010) Assembly algorithms for next-generation sequencing data. Genomics 95:315–327

Mitchelson KR (2007) New high throughput technologies for DNA sequencing and genomics. In:

Mitcheison KR (eds), Elsevier, Amsterdam, Netherlands, pp 303–326

Mullins LJ, Mullins JJ (2004) Insights from the rat genome sequence. Genome Biol 5(5):221

Nagarajan N, Pop M (2013) Sequence assembly demystified. Nat Rev Genet 14:157–167

Reinhardt J, Baltrus D, Marc N, Jeck W, Jones C, Dangl J (2008) De novo assembly using low-coverage short read sequence data from the rice pathogen *Pseudomonas syringae* pv *Oryzae*. Genome Res 19:294–305

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SIM et al (2009) ABySS: a parallel assembler for short read sequence data. Genome Res 19:1117–1123

Sohn J, Nam JW (2018) The present and future of *de novo* whole-genome assembly. Brief Bioinformat 19(1):23–40

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ (2001) The sequence of the human genome. Science 291(5507):1304–1351

Voelkerding KV, Dames SA, Durtschi JD (2009) Next-generation sequencing: from basic research to diagnostics. Clin Chem 55(4):641–658

Wang L, Yu S, Tong C, Zhao Y, Liu Y et al (2014) Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. Genome Biol 15 (2):R39

Warren RL, Sutton GG, Jones SJM, Holt RA (2007) Assembling millions of short DNA sequences using SSAKE. Bioinformatics 23(4):500–501

Wojcieszek M, Pawełkowicz M, Nowak R, Przybecki Z (2014) Genomes correction and assembling: present methods and tools. In: Photonics applications in astronomy, communications, industry, and high-energy physics experiments. International Society for Optics and Photonics 9290:92901X

Ye C, Hill CM, Wu S, Ruan J, Ma Z (2016) DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. Sci Rep 6:31900

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829

Zhang H, Miao H, Wang L, Qu L, Liu H et al (2013) Genome sequencing of the important oilseed crop *Sesamum indicum* L. Genome Biol 14(1):401

Zhao R, Miao H, Song W, Chen C, Zhang H (2018) Identification of sesame (*Sesamum indicum* L.) chromosomes using the BAC- FISH system. Plant Biol 20:85–92

Repetitive Sequences in Sesame Genome

14

Hongmei Miao, Yamin Sun, Wenchao Lin, Lei Wang, and Haiyang Zhang

Abstract

Repetitive DNA sequences are the homologous DNA fragments with multiple copies and are the main components in higher plant genomes. Reliable chromosome-scaled genome assembly provided abundant genome information regarding genome structure research in sesame. In this chapter, we present the genome structure character and distribution of interspersed repeat sequences and the tandem repeats in the assembled genome of sesame var. Yuzhi 11 (version 3.0). For *S. indicum*, the content of interspersed repeat sequences is 46.11%. Of the five groups of interspersed repetitive sequences, i.e., SINEs, LINEs, long terminal repeat (LTR) element, DNA element, and unknown dispersed repeat elements, LTR group is the biggest known repeat type and occupies 11.41% of the assembled genome (335 Mb). Ty1-*Copia* is the major superfamily in LTR-RTs. In sesame genome, distribution of the tandem repeats, such as rDNA, telomere, and centromere repeats, reflects the character of species. Cytogenetic analyses and genome structure showed that the individual 26 chromosomes in *S. indicum* have telomeric repeats at the terminal position. The telomeres of sesame are conserved, and the repeat sequences are same with the common repeat unit of TTTAGGG. A total of 1235 telomeric repeat of $(TTTAGGG)_3$ were found in the assembled genome. Meanwhile, about three thousand 45S and 5S rDNA repeats are detected in the assemble genome, which accord with the rDNA content range in other plants. Distribution characters and copy number of telomeres and rDNAs reflect the relative completeness of the assembled sesame genome. In addition, distribution of simple sequence repeats (SSRs) in sesame genome also has been precisely analyzed. A 153 bp centromeric repeat sequence is identified in sesame genome for the first time, which supplies more information for exploring the genome evolution of *Sesamum*.

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

Y. Sun · W. Lin · L. Wang
TEDA School of Biological Sciences and Biotechnology, Nankai University, Tianjin, P. R. China
e-mail: nksunyamin@aliyun.com

W. Lin
e-mail: linwenchao@yeah.net

L. Wang
e-mail: wanglei@nankai.edu.cn

H. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-98098-0_14

14.1 Introduction

Repetitive DNA sequences refer to the homologous DNA fragments with multiple copies in the genome (Mehrotra and Goyal 2014). In higher plants, repetitive sequences can occupy up to 90% of the nuclear genome (Mehrotra and Goyal 2014). Repetitive DNA sequences can be grouped into two major categories according to their genomic organization and chromosomal localization (Kubis et al. 1998). One category refers to the tandem repeating units, including satellite DNAs, telomeric repeats, and ribosomal DNA (rDNA). Copies of the repeated motifs are arranged adjacently to each other and preferentially locate at specific positions (such as chromosomal centromere and termini) on chromosomes. The other category refers to the interspersed repeat sequences such as short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs), and transposable elements (TEs). These repeat sequences are commonly but unevenly distributed throughout the genome (Kubis et al. 1998; Pisupati et al. 2018).

In earlier stage, the repetitive sequences have been regarded as the 'selfish elements' or 'junk DNA.' Subsequent studies proved that the type and content of the repetitive sequences vary in different genomes and might play important roles in genome evolution (Pisupati et al. 2018). Meanwhile, repetitive sequences contribute to genome size variation, mutations, chromosomal rearrangements, and gene regulation alteration (Gebre et al. 2016; Mascagni et al. 2018). Thus, to reveal the organization and the characteristics of sesame genome structure, the Sesame Genome Working Group (SGWG) screened all the repetitive sequences of the sesame genome and analyzed the content and location of major repetitive DNA elements. Distribution of main types of long terminal repeat retrotransposons (LTR-RTs) is described in this section. Meanwhile, the characters of telomeric repeats, 45S and 5S rDNAs, centromere repeats, and the microsatellite repeats or simple sequence repeats (SSRs) are also enumerated so as to elucidate the chromosomal evolution of the *Sesamum* species.

14.2 Interspersed Repetitive Sequences in Sesame Genome

14.2.1 Distribution of the Interspersed Repetitive Sequences

In angiosperm plants, the genome size (per haploid genome) obviously varies from ~61 megabase (Mb) in *Genlisea tuberosa* (Fleischmann et al. 2014; Tran et al. 2016) to 150 gigabase (Gb) in *Paris japonica* (Pellicer et al. 2010). Aside from polyploidy, the activity of transposable elements (TEs) is the main process contributing to the genome size variation in plants (Mascagni et al. 2018). TEs include two classes: class I and class II. Class I refers to the retrotransposons (mainly as LTRs), of which the retrotranscription and the integrity into the genome processes are dependent on the transcription of an RNA intermediate; class II refers to the DNA transposons (Magscagni et al. 2018). Insertion profile of the LTRs reflects the evolutionary character of the host genome in recent evolution history.

The SGWG initially analyzed the repetitive sequence composition in the chromosome-scaled genome assembly of sesame var. Yuzhi 11 (version 3.0) (Unpublished data, Haiyang Zhang). The interspersed repetitive sequences are screened using RepeatMasker (version 4.0.6) and Repbase database (version 20,150,807) (Table 14.1). There are five groups of interspersed repetitive sequences, i.e., SINEs, LINEs, long terminal repeat (LTR) element, DNA element, and unknown dispersed repeat elements. Results showed that the content of interspersed repeat sequences is 46.11%. Of the five groups of repetitive sequences, i.e., SINEs, LINEs, long terminal repeat (LTR) element, DNA element, and unknown dispersed repeat elements, LTR

group is the biggest known repeat type with the number of 42,529 and occupies 11.41% of the assembled genome (335 Mb). The second group is LINEs (6.67% of the assembled genome), in which LINE1 occupies 6.49%.

In flowering plants, the genome size differs by 2500-fold mainly for the variation of the copy number of TEs and other repetitive sequences (Pisupati et al. 2018). The repetitive elements in plant genomes vary from about 10% (in *Arabidopsis thaliana*) (Initiative and Copenhaver 2000) to more than 83% (in barley) (The International Barley Genome Sequencing Consortium 2012). Compared with several genomes close to *S. indicum* in the phylogenetic tree of angiosperm plants, such as potato (*Solanum tuberosum*) (62.2% of the assembled 760 Mb genome) (Xun et al. 2011), cacao (*Theobroma cacao*) (25.7% of the assembled 326.9 Mb genome) (Argout et al. 2011), and grapevine (*Vitis vinifera*) (41.4% of the assembled 487 Mb genome) (Jaillon et al. 2007), the content of the repetitive/transposable elements (46.11%) of sesame genome is moderate.

14.2.2 LTR Components and LTR Insertion

In higher plants, the structure of full-length LTR-RTs comprises 5' LTR, 3' LTR, and the retrotransposons. LTRs vary in size from 1 to 5 kb (Wessler 1996). In Table 14.1, LTR elements are the main known repeat type in sesame genome and occupy 11.4% of the whole genome. Considering that the dynamics of the LTR-RTs reflect the architecture of the eukaryotic genomes, the SGWG analyzed the distribution and the characters of LTR-RTs in sesame genome using LTR-FINDER software (version 1.06) (Xu and Wang 2007). As a result, 8208 LTR-RTs are detected in sesame (Table 14.2) (Unpublished

Table 14.1 Distribution of repetitive sequences in sesame genome*

TE type	TE number	Size (bp)	Percentage of genome (%)
SINEs	697	536,672	0.16
ALUs	0	0	0.00
MIRs	0	0	0.00
Others	697	536,672	0.16
LINEs	19,020	22,373,497	6.67
LINE1	16,520	21,764,990	6.49
LINE2	0	0	0.00
L3/CR1	1382	392,200	0.12
Others	1118	216,307	0.06
LTR elements	42,529	38,242,140	11.41
Copia	24,985	18,056,639	5.39
Gypsy	16,366	15,009,404	4.48
Others	1178	5,176,097	1.54
DNA elements	27,807	10,869,502	3.24
hAT-Charlie	97	3976	0.00
TcMar-Tigger	0	0	0.00
Others	27,710	10,865,526	3.24
Unclassified	261,846	82,543,222	24.63
Total interspersed repeats	351,899	154,565,033	46.11

*Provided by Haiyang Zhang

data, Haiyang Zhang). The results showed that all the LTR-RTs are comprised of the specific LTR-RT elements: terminal TG/CA inverted repeats in the LTRs and the target site duplications (TSDs) of 4–6 bp (Cossu et al. 2012; Nystedt et al. 2013), of which 1086 (13.23%) are putative full-length LTR-RTs and 57,122 (86.77%) are LTR-RT remnants. In sesame genome, the LTR-RT remnants present the high ratio phenomenon similar to other plants (Bennetzen et al. 2004; Natali et al. 2015).

Further analysis indicated that all these 1086 LTR-RTs are 8.5 Mb in length with the average length of 7785 bp and occupy 2.54% of the genome size. In sesame genome, a total of 912 (11.11%) LTR-RTs dispersed all the 13 chromosomes (Fig. 14.1). The average number of LTR-RTs per chromosome varies from 50 (in *Si*Chr.13) to 98 (*Si*Chr.1). For the 10 chromosomes from *Si*Chr.1 to *Si*Chr.10, the high density of LTR-RTs is mainly present at the central or near central position of chromosomes.

To explore the LTR amplification and the genome DNA loss, we calculated the LTR insertion time based on all the full-length LTR-RTs in sesame (data not shown). The nucleotide distance of LTR-RTs was estimated using the Kimura two-parameter (K2P) (transition–

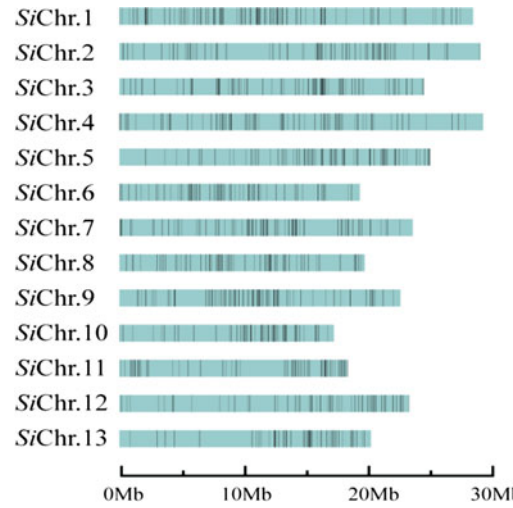


Fig. 14.1 Distribution diagram of the putative full-length LTR-RTs in sesame genome. A total of 912 of 1086 putative full-length LTR-RTs are localized on the 13 chromosomes in sesame. Horizontal line refers to the genome sequence (Mb). The vertical black bar refers to the LTR retrotransposon

Table 14.2 Distribution of LTR-RTs in *S. indicum* genome

Chromosome no.	LTR-RT number	LTR-RT length (bp)	Percentage of genome (%)
*Si*Chr.1	1148	2,097,220	0.63
*Si*Chr.2	107	636,588	0.19
*Si*Chr.3	166	759,557	0.23
*Si*Chr.4	113	692,168	0.21
*Si*Chr.5	163	826,052	0.25
*Si*Chr.6	176	687,569	0.21
*Si*Chr.7	135	762,432	0.23
*Si*Chr.8	2852	939,621	0.28
*Si*Chr.9	104	709,617	0.21
*Si*Chr.10	135	638,352	0.19
*Si*Chr.11	215	803,593	0.24
*Si*Chr.12	184	795,612	0.24
*Si*Chr.13	811	955,892	0.29
Other scaffolds	1899	5,451,109	1.63
Total	8208	16,755,382	5.03

Note Other scaffolds refer to the scaffolds unanchored onto the 13 chromosomes. All the data are provided by Haiyang Zhang

transversion ratio) criterion (Kimura 1980). The insertion time (T) of LTR-RTs was calculated as T = D/2μ, with μ as the rate of nucleotide substitution. Considering the genome size similarity and the annual growth character, the mutation rate (μ) was set at 1.3×10^{-8} substitutions per synonymous site per generation or year as that of *Oryza glaberrima* (Ma and Bennetzen 2004). ClustalW analysis of 1,086 LTR-RTs showed that there is no obvious insertion of LTR. Compared to the mean insertion age of LTR-RTs in maize (~1 Myr, with mutation rate of 1.3×10^{-8} substitutions/site/year) (Brunner et al. 2005), *Arabidopsis thaliana* (3.1 Myr, with mutation rate of 7×10^{-9}) (Hu et al. 2011), *Sorghum bicolor* (average 0.8 Myr, with mutation rate of 1.3×10^{-8}) (Paterson et al. 2009), wheat (<3 Myr, with mutation rate of 1.3×10^{-8}) (Charles et al. 2008), and sunflower (0–1.4 Myr, 1.0×10^{-8}) (Staton et al. 2012), LTR-RTs of sesame are generally older than those of the above annual species (Unpublished data, Haiyang Zhang).

According to the insertion time of each 912 full-length LTR retrotransposons, the SGWG explored and drew the LTR-RT distribution in various chromosome regions (Fig. 14.2) (Unpublished data, Haiyang Zhang). Each chromosome showed the variation of insertion age with various LTR-RTs. The LTR-RTs in the regions above the mean insertion time line (1.5 Myr) should expand in earlier era.

Plant LTR-RTs are primarily divided into two main superfamilies, i.e., Ty1-*Copia* and Ty3-*Gypsy*, based on the order of POL protein domains (protease, retrotranscriptase, integrase, and RNAseH) and on sequence similarity (Wicker et al. 2007; Natali et al. 2015). Previous studies showed that LTR-RT families presented the different amplification rates in various genera or species (Magscagni et al. 2018). In different genera, the ratio of *copia* and *gypsy* elements varies from 5:1 (in papaya) (Ming et al. 2008) to 1:2 (in grapevine) (Jaillon et al. 2007). Therefore, the SGWG group further screened the *copia* and *gypsy* elements in sesame (Table 14.3) (Unpublished data, Haiyang Zhang). A total of 24,985 *copia* and 16,366 *gypsy* elements were found. The ratio of *copia* and *gypsy* was 1.5:1; 94.24% copia and 94.62% gypsy were distributed onto the 13 chromosomes.

Subsequently, the SGWG differentiated the full-length *copia* and *gypsy* elements in the 1086 full-length LTR retrotransposons in sesame genome (Unpublished data, Haiyang Zhang). Apart from the 2865 other or unknown elements, the majority of retrotransposons are *gypsy* subfamily (1108) (Table 14.4). The ratio of *gypsy* and *copia* elements within the 1086 LTR retrotransposons is 1.5:1. The results indicated that the activities of *gypsy* and *copia* elements are similar to each other (data not shown). Distribution of the *copia* and *gypsy* elements in the sesame genome reflects the characters of LTR-RTs.

14.3 Simple Sequence Repeats

SSRs exhibit high mutation rates and are also abundant within plant genomes. Therefore, the SGWG screened the genome using the home-made script and detected the SSRs in sesame (Table 14.5) (Unpublished data, Haiyang Zhang). A total of 153,488 perfect and compound SSRs were detected in sesame genome. The total length of the SSRs occupied 0.96% of the genome size (335 Mb), of which 119,557 (77.89%) are the SSRs with more than 10 nucleotide repeats. For the perfect SSRs ($\geq$15 bp), six types with the repeat motifs ranging from 1 (mono-) to 6 (hexa-) nucleotides are found in sesame (Table 14.5). Di-motif type occupies 50.83% and is the first main SSR repeats in sesame, followed by mono-motif repeat (27.54%) and tri-motif repeat (19.03%). All the SSRs distributed onto all parts of 13 chromosomes. *Si*Chr.2 carries the highest number of SSRs, while SiChr.7 has the high SSR density of 738 bp per SNP.

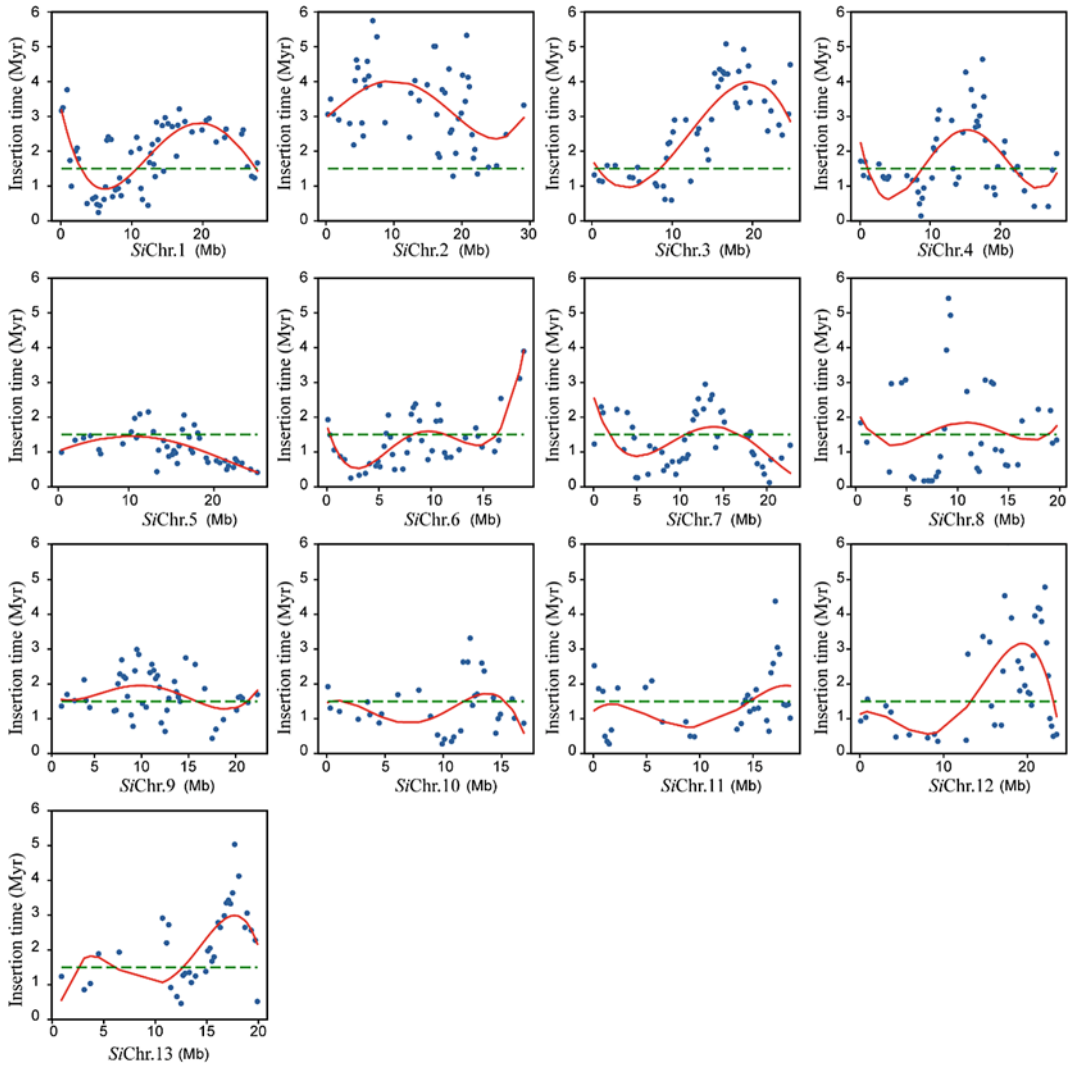


Fig. 14.2 Mean insertion time of full-length LTR-RTs in the assembled genome of *S. indicum*. X-axis indicates the genome sequence with a scan window of 200 kb. Y-axis indicates the insertion time (Myr). Red smoothed curve is obtained based on the average values of each three neighboring LTR-RTs. Green dashed line is the mean insertion time of sesame LTR-RTs (1.5 Myr). All the data are supplied by Haiyang Zhang

14.4 Telomeres and rDNA

14.4.1 Telomeres

Telomeres are the specific structure of the termini of linear eukaryotic chromosomes (Richards and Frederick 1988). Telomeres are composed of the high conserved minisatellite DNA sequences which can protect the completeness of the chromosomes. In most plants, telomeric repeat is (TTTAGGG)n, and the repeat copy number is high to 20,000 repeats (Fuchs et al. 1995; Lee et al. 2016). For example, in *Arabidopsis*, the telomeres are composed of (CCCTAAA) repeats with 2–3 kb in length (Richards and Ausubel 1988; Lee et al. 2016). In order to clarify the telomere repeat sequence and distribution in genome, the SGWG-HSRC firstly performed fluorescence in situ hybridization (FISH) on

Table 14.3 Distribution of main LTR retrotransposons in sesame genome

Chromosome no.	Copia number	Copia length (bp)	Gypsy number	Gypsy length (bp)
*Si*Chr.1	1463	952,423	1030	927,797
*Si*Chr.2	1199	874,006	1290	1,076,028
*Si*Chr.3	1308	871,508	1124	1,012,831
*Si*Chr.4	2164	1,550,941	1246	1,052,335
*Si*Chr.5	1789	1,230,430	1634	1,445,483
*Si*Chr.6	1820	1,402,784	1019	882,921
*Si*Chr.7	1631	1,233,824	1426	1,178,380
*Si*Chr.8	2079	1,464,938	1827	1,497,289
*Si*Chr.9	2023	1,411,451	1630	1,346,769
*Si*Chr.10	1273	987,420	1097	801,154
*Si*Chr.11	2086	1,619,868	1432	1,252,471
*Si*Chr.12	1464	1,032,792	1111	861,317
*Si*Chr.13	2007	14,01,837	1297	1,115,230
Other scaffolds	1438	1,983,592	880	1,353,619
Total	24,985	18,056,639	16,366	15,009,404

Note Other scaffolds refer to the scaffolds unanchored onto the 13 chromosomes. Supplied by Haiyang Zhang

Table 14.4 Distribution of the full-length copia and gypsy elements in 1086 full-length sesame LTR retrotransposons

Chromosome no.	Full-length LTR retrotransposon			
	Copia number	Gypsy number	Other element	Total LTR
*Si*Chr.1	51	76	441	98
*Si*Chr.2	60	71	98	73
*Si*Chr.3	65	74	126	77
*Si*Chr.4	90	83	137	82
*Si*Chr.5	48	86	222	75
*Si*Chr.6	60	77	129	66
*Si*Chr.7	65	98	165	79
*Si*Chr.8	64	57	128	61
*Si*Chr.9	50	123	229	78
*Si*Chr.10	53	86	92	57
*Si*Chr.11	31	99	156	58
*Si*Chr.12	58	55	142	58
*Si*Chr.13	32	82	175	50
Other scaffolds	13	41	65	174
Total	740	1108	2865	1086

Note Other scaffolds refer to the scaffolds unanchored onto the 13 chromosomes

sesame metaphase chromosomes using the universal plant telomeric repeat probe (5'-TTTAGGGTTTAGGGTTTAGGG-3') labeled with fluorescent group FAM (Fig. 14.4a–c) for the first time (Unpublished data, Haiyang Zhang).

Table 14.5 Number and distribution of perfect SSR repeats (SSR $\geq$ 15 bp) in sesame genome

Chromosome no.	SSR repeat number						
	Mono-	Di-	Tri-	Tetra-	Penta-	Hexa-	Total
*Si*Chr.1	884	1571	588	63	11	10	3127
*Si*Chr.2	1046	1869	594	70	11	17	3607
*Si*Chr.3	864	1754	640	62	7	19	3346
*Si*Chr.4	951	1636	599	58	14	11	3269
*Si*Chr.5	740	1517	501	63	8	10	2839
*Si*Chr.6	703	1411	461	46	9	16	2646
*Si*Chr.7	691	1296	435	45	5	11	2483
*Si*Chr.8	652	1158	369	48	1	9	2237
*Si*Chr.9	626	1340	503	50	18	8	2545
*Si*Chr.10	554	1205	392	39	6	7	2203
*Si*Chr.11	557	1035	414	36	11	9	2062
*Si*Chr.12	643	1056	421	50	3	5	2178
*Si*Chr.13	583	1152	425	35	9	12	2216
Other scaffolds	422	302	510	12	1	2	1249
Total	9916	18,302	6852	677	114	146	36,007

Note Other scaffolds refer to all the scaffolds unanchored onto the 13 chromosomes. All the data are supplied by Haiyang Zhang

As shown in Fig. 14.4a–c, green color signal of the common telomeric repeat probe in plant was observed at the ends of each sesame chromosome. The results indicated that the specific DNA regions at the termini of the 13 pairs of chromosomes hybridized by the telomeric repeat probe have telomere repeat sequences. For sesame, the telomere repeats are conserved and same with the common repeat unit sequence of TTTAGGG. Thus, the SGWG-HSRC screened the telomere repeats in the assembled genome (Unpublished data, Haiyang Zhang).

14.4.2 Ribosomal DNA

Ribosomal DNA (rDNA) repeat sequences comprise of 45S (18S-5.8S-25S) and 5S rDNA sequences. Similar to telomere repeats, rDNA repeats also locate at specific regions on chromosomes. Ribosomal DNA sequences are usually specific in various species and reflect the architecture of the chromosome sets and the specific genomes. In order to determine the physical position of the rDNA sequences in sesame chromosome set, 45S and 5S rDNA repeat sequences of *Arabidopsis* were initially labeled and applied in FISH system of sesame by Zhao et al. (2018). The 45S rDNA repeats include 5'-GAA TTG CAG AAT CC-3' (5.8S), 5'-CGC TGC GTT CTT CAT CG-3' (5.8S), 5'-GTA GTC ATA TGCTTGTCTC-3' (18S), 5'-TGG CAC CAG ACT TGC CCT-3' (18S), 5'-TCG TAA CAA GGT TTC CGT AG-3' (18S), 5'-CGA CCC CAG GTC AGG CG-3' (26S), and 5'-GAG TCG GGT TGT TTG GGA-3' (26S), while 5S rDNA repeat sequence includes 5'-CTGATGGGATCCGGTGCTTT-3'. During hybridization, the probes are labeled with the tetramethyl-rhodamine and the fluorescent group FAM to generate red and green color signals, respectively.

As shown in Fig. 14.3d–i, three pairs of satellite chromosomes carrying 45 rDNA repeats (green signal) and two pairs of chromosomes with 5S rDNA (red signal) were observed in *S. indicum*. According to the nomenclature of sesame chromosome set described by Zhao et al.

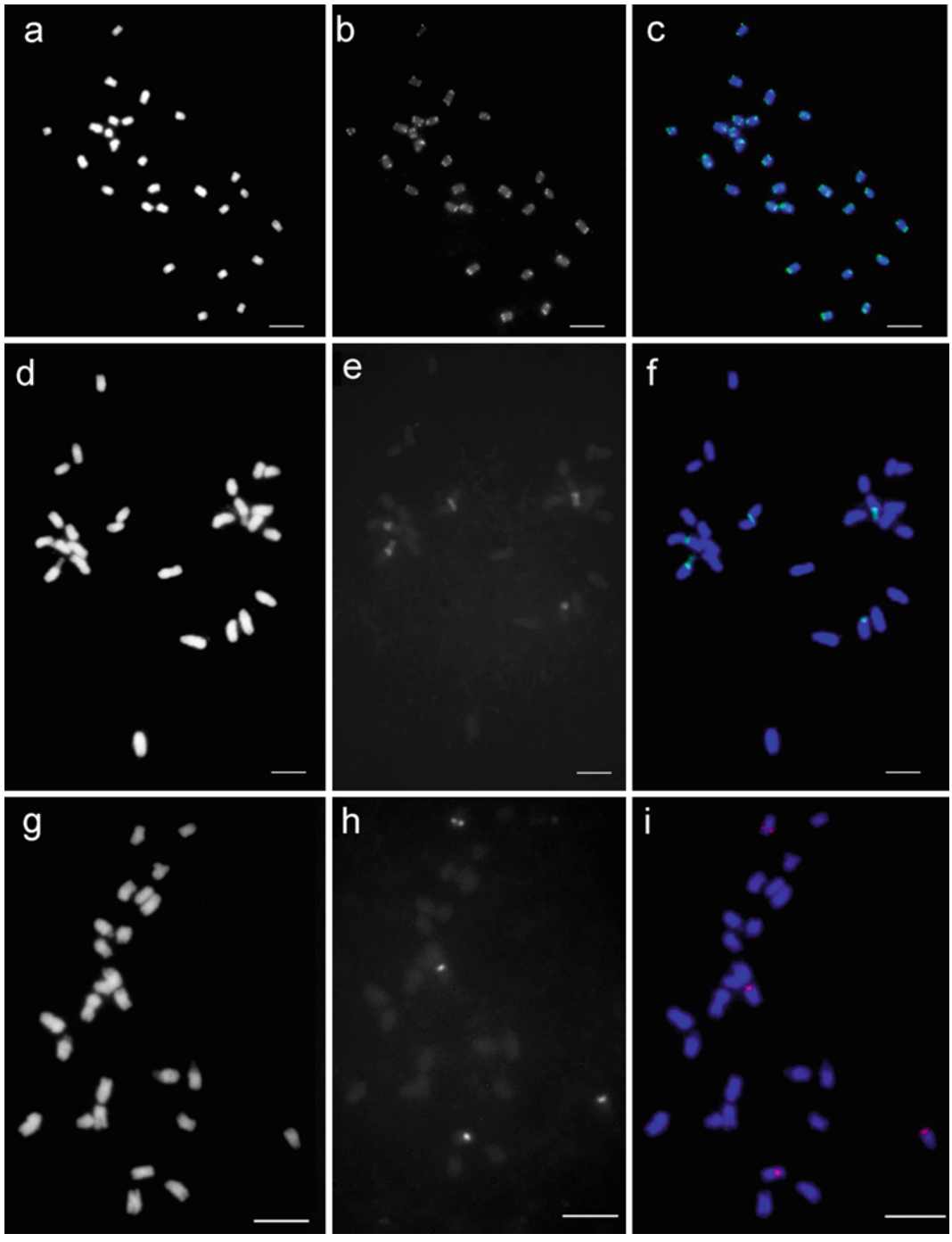


Fig. 14.3 Fluorescence in situ hybridization of sesame metaphase chromosomes using telomeric repeat and rDNA probes. **a**, **d**, **g** Chromosomes of var. Yuzhi 11 dyed by DAPI. **b**, **e**, **h** Chromosomes hybridized using telomeric repeat, 45S rDNA, and 5S rDNA probes, respectively. **c**, **f**, **i** Image integrity of chromosomes hybridized by telomeric repeat, 45S rDNA, and 5S rDNA probes, respectively. The red and green signals are released from the probes labeled with tetramethyl-rhodamine and fluorescein, respectively. Bar = 5 μm (Supplied by Haiyang Zhang)

(2018), 45S rDNA sequences are located on *Si*Chr.11, *Si*Chr.12, and *Si*Chr.13, respectively, while 5S rDNA sequences are located on *Si*Chr.6 and *Si*Chr.7, respectively, in the chromosome set of *S. indicum* (described in Chap. 9).

Compared with 45S rDNA, 5S rDNA sequences are not always located at chromosome terminal regions in sesame. These results suggest that 45S and 5S rDNAs do not tend to integrate in chromosomes. They might undergo different evolutionary processes to yield distinct distribution patterns (Tagashira and Kondo 2001). Thus, to illustrate the physical mapping of telomeres and rDNAs, we performed further genomics analysis in sesame genome.

14.4.3 Distribution of Telomeres and rDNAs in Sesame Genome

To explore the distribution character and the genome structure of telomeres and rDNAs in sesame, the SGWG-HSRC further analyzed the copy and physical position of telomeres, 45S rDNAs, and 5S rDNAs, respectively, in the assembled sesame genome (Table 14.6) (Unpublished data, H. Zhang). The threshold of the nucleotide identity for telomeres, 5S rDNA, and 45S rDNA was set at 100%, 100%, and 90%, respectively. As a result, a total of 1235 telomeric repeat of $(TTTAGGG)_3$ were found in the assembled genome, of which 51.17% of telomeric repeat $(TTTAGGG)_3$ was distributed onto the 13 chromosomes. The other 48.83% repeats were located in the 14 scaffolds which have not been assembled onto the chromosome set for the high repetitive ratio. The $(TTTAGGG)_3$ repeat copies of the 13 chromosomes varied from 2 (on *Si*Chr.6) to 142 (on *Si*Chr.2).

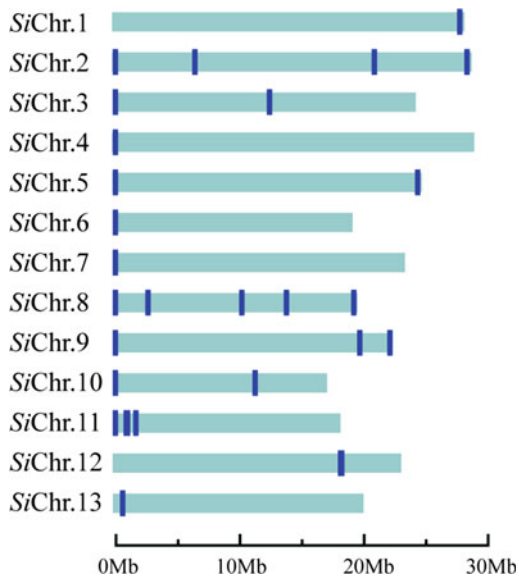


Fig. 14.4 Distribution diagram of the telomeric repeats in sesame genome. Blue bar indicates telomeric repeats $(TTTAGGG)_n$ in sesame genome. Green segment indicates the pseudochromsome molecule of sesame genome

Further genome screening analysis indicated that there are 26 telomeric repeat blocks located on the sequences of the 13 chromosomes (Fig. 14.4), of which 16 repeat blocks are localized at the chromosomal terminals. The copy number of the assembled telomeric repeat $(TTTAGGG)_3$ per telomeric repeat block varies from 1 (21 bp) to 113 (about 2.4 kb). For the 8 normal chromosomes which differ from the 3 satellite chromosomes and 2 specific chromosomes with 5S rDNA, 4 chromosomes (i.e., *Si*Chr.2, 5, 8, and 9) are proved assembled to the terminal position. For the 5 chromosomes (*Si*Chr.6, 7, 11, 12, and 13) carrying 45S or 5S rDNA sequences, the telomeric repeats are only present at single ends and some are opposed to the rDNA ends. In a total, distribution of telomeric repeats reflects the high completeness of the genome assembly for sesame.

Moreover, the length of the 16 repeat blocks of the telomeric repeats localized at the chromosomal terminals has been calculated, based on the assembled genome sequences (Unpublished data, Haiyang Zhang). As a result, the largest assembled telomere at the *Si*Chr.9 chromosomal terminals is 2373 bp, which is similar to the average telomeric repeat length (about 2–3 kb) of *Arabidopsis* (Initiative and Copenhaver 2000; Richards and Frederick 1988). The largest telomeric repeat spans 10 kb in sesame genome.

As to the 45S rDNA, a total of 2549 repeats related to 5.8S, 18S, and 26S rDNAs have been detected in the assembled sesame genome

Table 14.6 Distribution of telomeres and rDNA sequences in sesame genome

Chromosome no.	Repeat copy		
	Telomere repeat (n = 3)	45S rDNA repeat	5S rDNA repeat
*Si*Chr.1	16	0	0
*Si*Chr.2	142	0	0
*Si*Chr.3	19	5	1
*Si*Chr.4	22	15	0
*Si*Chr.5	118	9	0
*Si*Chr.6	2	8	388
*Si*Chr.7	17	12	0
*Si*Chr.8	13	0	1
*Si*Chr.9	115	0	0
*Si*Chr.10	27	0	0
*Si*Chr.11	109	33	0
*Si*Chr.12	15	40	0
*Si*Chr.13	17	71	0
Scaffold with repeats	603 (14)	2356 (349)	331 (2)
Total	1235	2549	721

Note Data in brackets indicate the number of the scaffolds with telomeric and rDNA repeats. Scaffold with repeats refers to the scaffold unanchored onto the 13 chromosomes. All the data are supplied by Haiyang Zhang

(Table 14.6). About 193 (7.57%) repeats are localized onto all the 8 chromosomes, while 2356 (92.43%) are distributed in the 349 scaffolds unassembled onto 13 chromosomes. Of the 193 45S rDNA repeats, 74.61% are localized onto three chromosomes of *Si*Chr.11 (17.10%), *Si*Chr.12 (20.72%), and *Si*Chr.13 (36.79%). The rest 25.4% are localized onto 7 chromosomes. Interestingly, most 45S rDNA repeats of *Si*Chr.11, 12, and 13 are determined at or near the end of chromosomes, even though the other 45S rDNA repeats (in Fig. 14.3d–f) are dispersed on chromosomes. Distribution of the 45S rDNA repeats assembled in *Si*Chr.11, 12, and 13 sequences is consistent with the FISH results of sesame chromosome set, which also proved the high quality of the sesame genome assembly.

For 5S rDNA, 390 (54.09%) and 331 (45.91%) repeats are detected on 6 chromosomes and 2 unassembled scaffolds, respectively. Especially, 388 (99.49%) of the 390 rDNA repeats are located on *Si*Chr.6 which belongs to the specific chromosomes carrying 5S rDNA in the cytogenetic map of sesame (described in Chap. 9). Further physical position in genome proved that the 5S rDNA repeat blocks cover 0.24 Mb and are close to the terminal position of *Si*Chr.6 (Fig. 14.3g–i) (Unpublished data, Haiyang Zhang). However, the scaffolds with 5S rDNA sequences are not successfully assembled into the genome for the high repetition level and might result in the deficiency of 5S rDNA repeats on the pseudochromsome molecule of *Si*Chr.7.

In general, plants have 500–40,000 rDNA copies per diploid cell. For sesame, about three thousand 45S and 5S rDNA repeats are detected at the terminal and near-terminal regions of chromosomes, based on the chromosome-scaled genome assembly, which is in accordance with the rDNA content characters in plants. We here combine the analysis results of the telomeres, 45S rDNA, and 5S rDNA in sesame genome for the first time and finally determine the copy and distribution characters of telomeres and 45S and 5S rDNAs in sesame. The findings of the three types of the repetitive repeats reflect the structure

of sesame genome and supply more information for genome and chromosome set evolution analysis of *Sesamum* species in the future.

14.5 Centromeres

14.5.1 Centromeric Repeats in Sesame

Centromere is the site of spindle fiber attachment and is responsible for chromosome movement at mitosis and meiosis (Henikoff et al. 2001). In most eukaryotes, centromere is comprised of highly repetitive centromeric retrotransposons and satellite repeats. However, the length of the centromeric repeats varies evidently in plants (Melters et al. 2013). For example, the centromere repeat of soybean is about 92 bp (Tek et al. 2010), while monkey flower has a 728 bp centromere repeat (Fishman and Sauders 2008). In addition, the centromeric DNA sequences are divergent even among close related species (Burrack and Berman 2012; Lermontova et al. 2015). Low conversation of the centromeric DNA in species might be the epigenetic control result, rather than the primary DNA sequence (Ugarković 2009), but increase the difficulty to exploring the genome structure and evolution characters of a species.

Previous studies proved that the centromeric repeats originated from retrotransposon-related sequences in *G. raimondii* (Han et al. 2016). Therefore, in order to determine the centromeres and the centromeric repeats or satellite DNAs in centromeres of sesame, the SGWG-HSRC firstly screened the 27 specific bacterial artificial chromosomes (BACs) from the sesame BAC library (var. Yuzhi 11) constructed by Henan Sesame Research Center, Henan Academy of Agricultural Sciences (HSRC, HAAS), for centromeric repeat screening (Fig. 14.5) (Unpublished data, Haiyang Zhang). The specific BAC-FISH technology constructed by the SGWG-HSRC (Zhao et al. 2018) was applied for the research. As the 3 BAC probes B428R, B429R, and B400R were successfully hybridized in centromeric and pericentromeric regions of the 13 chromosome pairs in sesame (Fig. 14.5c, f, i), the 3 BAC clones would carry centromeric or other repetitive DNA sequences. According to the image requirement, only the BACs as FISH probes hybridized on the same chromosomal region concentrated with DAPI could be chosen for further analysis. As a result, the SGWG-HSRC collected 27 BACs with 13 pairs of the specific BAC-FISH signals located at the centromeric and pericentromeric regions of all the 13 chromosome pairs.

Subsequently, the group analyzed the sequences of the 27 BACs using RepeatMasker (version 4.0.6) (Table 14.7) to detect LTR retrotransposons. Of the 27 BACs, 15 were detected with LTR retrotransposons. Then, all the LTR types in the 15 target BACs were annotated using RepeatMasker (version 4.0.6) to determine the species-specific repetitive sequences. Alignment of the 27 LTR sequences was also performed using MAFFT V7.245 (data not shown). As a result, a 153 bp repeat sequence was determined from the specific BACs and named '*Si*Cen1' as the candidate-specific centromere repeat in sesame. Further screening of *Si*Cen1 repeat was performed with the following threshold parameters: $P < 1e{-}5$, alignment length ≥ 120 bp, and identity ratio $\geq 90\%$. In sesame genome, 1328 copies of the *Si*Cen1 repeats are found which are concentrated at the central region of 12 chromosomes except for *Si*Chr.4 (Unpublished data, Haiyang Zhang), which indicates that the DNA sequences of most of the 13 pseudochromsomes for sesame are complete. The findings supply new information for centromere repeat detection and the genome evolution analysis in *Sesamum* genus.

14.5.2 Satellite DNA in Sesame

To reflect the distribution of satellite DNAs in sesame, the SGWG-HSRC also screened the satellite DNA repeats in the assembled genome using RepeatMasker software (4.0.6). The results showed that 1227 satellite DNA repeats exist in sesame and occupied 0.07% of the assembled genome sequences (Table 14.7). All the satellite DNA repeats distribute all parts of the 13

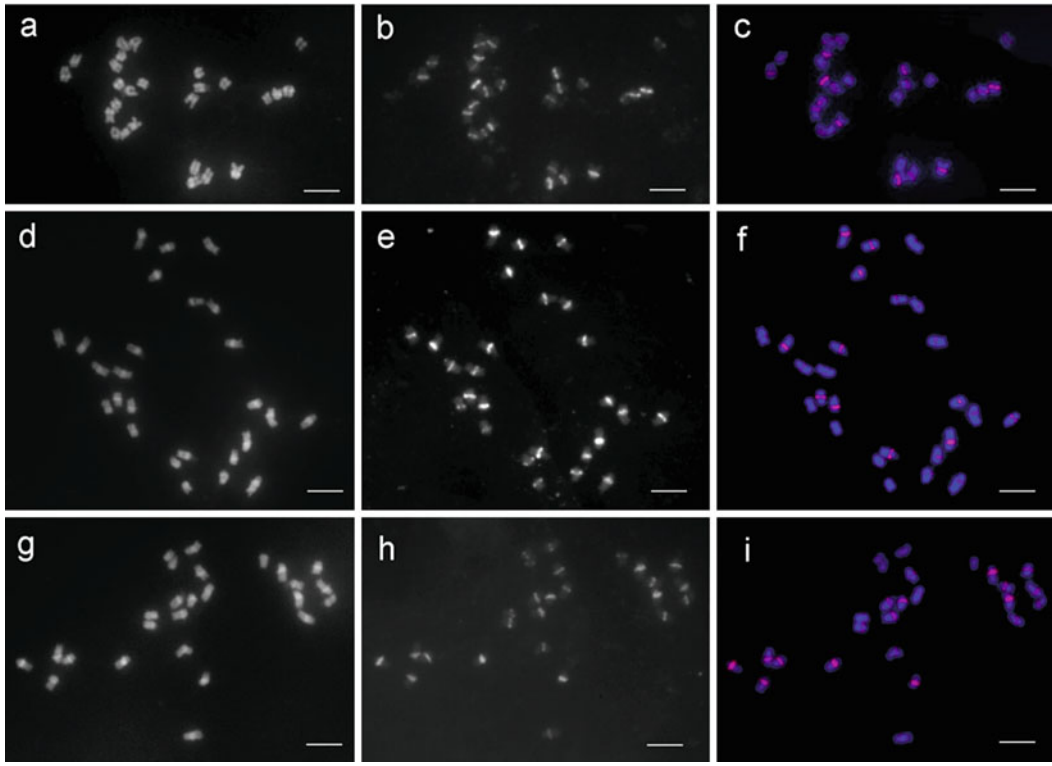


Fig. 14.5 Fluorescence in situ hybridization of BACs on sesame metaphase chromosomes. **a**, **d**, **g** Chromosomes of *var.* Yuzhi 11 dyed by DAPI. **b**, **e**, **h** Chromosomes hybridized using BAC B428R, B429R, and B400R probes, respectively. **c**, **f**, **i** Image integrity of chromosomes hybridized by B429R and B400R probes, respectively. The red signals are generated from the probes labeled with tetramethyl-rhodamine and fluorescein, respectively. The specific chromosome position hybridized by the above BACs is the same with the region concentrated with DAPI. Bar = 5 μm (Provided by Haiyang Zhang)

Table 14.7 Distribution of *Si*Cen1 repeat sequences in sesame genome

Chromosome no.	*Si*Cen1 repeat copy	Satellite DNA repeat copy	Satellite DNA length (bp)
*Si*Chr.1	618	114	23,759
*Si*Chr.2	20	121	19,723
*Si*Chr.3	51	104	18,390
*Si*Chr.4	0	108	21,231
*Si*Chr.5	93	84	15,785
*Si*Chr.6	203	111	22,898
*Si*Chr.7	66	102	25,170
*Si*Chr.8	18	85	18,224
*Si*Chr.9	100	87	19,690
*Si*Chr.10	23	85	20,542
*Si*Chr.11	6	67	15,365
*Si*Chr.12	56	80	14,861
*Si*Chr.13	74	68	14,805
Other scaffolds	0	11	2421
Total	1328	1227	252,864

Note Data in brackets indicate the number of the scaffolds with telomeric and rDNA repeats. Other scaffolds refer to the scaffolds unassembled onto the 13 chromosomes (Provided by Haiyang Zhang)

chromosomes, even though the distribution is not even (data not shown). The average length of the satellite DNA in sesame is 206 bp, relatively in accordance with the range of 100–200 bp in many plant species (Melters et al. 2013). The number of satellite DNA repeats in individual chromosomes varies from 67 (in *Si*Chr.11) to 121 (in *Si*Chr.2). Compared to the centromeric repeats, the positions with high-density satellite DNA repeats seem overlapped with the centromere regions and reflect the high identity relationship between centromeric repeats and satellite DNA repeats in sesame chromosome set (data not shown). All the information supplies a solid foundation for further genomic evolution analysis of the *Sesamum* species. On the other hand, the above results prove the high quality of the genome assembly of *S. indicum* (var. Yuzhi 11). Chromosome site-specific repetitive sequences generally show a high rate of nucleotide substitution and are divergent between even closely related species (Kaszás and Birchler 1998; Donget al. 1998). Therefore, the SGWG applies the reliable genome assembly and the abundant genome information in genome synteny and chromosome set rearrangement in *Sesamum* species (partially shown in the following chapters).

References

Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G et al (2011) The genome of *Theobroma cacao*. Nat Genet 43(2):101–108

Bennetzen JL, Coleman C, Liu R, Ma J, Ramakrishna W (2004) Consistent over-estimation of gene number in complex plant genomes. Curr Opin Plant Biol 7 (6):732–736

Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A (2005) Evolution of DNA sequence nonhomologies among maize inbreds. Plant Cell 17(2):343–360

Burrack LS, Berman J (2012) Neocentromeres and epigenetically inherited features of centromeres. Chrom Res 20(5):607–619

Charles M, Belcram H, Just J, Huneau C, Viollet A et al (2008) Dynamics and differential proliferation of transposable elements during the evolution of the B and A genomes of wheat. Genetics 180(2):1071–1086

Cossu RM, Buti M, Giordani T, Natali L, Cavallini A et al (2012) A computational study of the dynamics of LTR retrotransposons in the *Populus trichocarpa* genome. Tree Genet Genom 8(1):61–75

Dong F, Miller JT, Jackson SA, Wang GL, Jiang JM (1998) Rice (*Oryza sativa*) centromeric regions consist of complex DNA. Proc Natl Acad Sci USA 95 (14):8135–8140

Fishman L, Saunders A (2008) Centromere-associated female meiotic drive entails male fitness costs in monkey flowers. Science 322(5907):1559–1562

Fleischmann A, Michael TP, Rivadavia F, Sousa A, Wang W (2014) Evolution of genome size and chromosome number in the carnivorous plant genus *Genlisea* (Lentibulariaceae), with a new estimate of the minimum genome size in angiosperms. Ann Bot 114(8):1651–1663

Fuchs J, Brandes A, Schubert I (1995) Telomere sequence localization and karyotype evolution in higher plants. Plant Syst Evol 196(3–4):227–241

Gebre YG, Bertolini E, Pè ME, Zuccolo A (2016) Identification and characterization of abundant repetitive sequences in *Eragrostis tef* cv Enatite genome. BMC Plant Biol 16(1):39

Han J, Masonbrink RE, Shan W, Song F, Zhang J et al (2016) Rapid proliferation and nucleolar organizer targeting centromeric retrotransposons in cotton. Plant J 88(6):992–1005

Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293(5532):1098–1102

Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF et al (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat Genet 43 (5):476

Initiative TAG, Copenhaver GP (2000) Arabidopsis genome initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408.6814 (2000):796

Jaillon O, Aury JM, Noel B, Polocriti A, Clepet C et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449(7161):463

Kaszás E, Birchler JA (1998) Meiotic transmission rates correlate with physical features of rearranged centromeres in maize. Genetics 150(4):1683–1692

Kimura M (1980) Simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kubis S, Schmidt T, Heslop-Harrison JS (1998) Repetitive DNA elements as a major component of plant genomes. Ann Bot 82(suppl_1):45–55

Lee JR, Xie X, Yang K, Zhang J, Lee SY et al (2016) Dynamic interactions of Arabidopsis TEN1: stabilizing telomeres in response to heat stress. Plant Cell 28:2212–2224

Lermontova I, Sandmann M, Mascher M, Schmit AC, Chabouté ME (2015) Centromeric chromatin and its dynamics in plants. Plant J 83(1):4–17

Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. Proc Natl Acad Sci USA 101(34):12404–12410

Mascagni F, Usai G, Natali L, Cavallini A, Giordani T (2018) A comparison of methods for LTR-retrotransposon insertion time profiling in the *Populus trichocarpa* genome. Caryologia 71(1):85–92

Mehrotra S, Goyal V (2014) Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. Genom Proteom Bioinformat 12(4):164–171

Melters DP, Bradnam KR, Young HA, Telis N, May MR et al (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol 14(1):1–20

Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A et al (2008) The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). Nature 452(7190):991

Natali L, Cossu RM, Mascagni F, Giordani T, Cavallini A et al (2015) A survey of Gypsy and Copia LTR-retrotransposon superfamilies and lineages and their distinct dynamics in the *Populus trichocarpa* (L.) genome. Tree Genet Genom 11(5):107

Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y et al (2013) The Norway spruce genome sequence and conifer genome evolution. Nature 497(7451):579

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457(7229):551

Pellicer J, Fay MF, Leitch IJ (2010) The largest eukaryotic genome of them all? Bot J Linn Soc 164(1):10–15

Pisupati R, Vergara D, Kane NC (2018) Diversity and evolution of the repetitive genomic content in *Cannabis sativa*. BMC Genom 19(1):156

Richards EJ, Ausubel FM (1988) Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. Cell 53(1):127–136

Staton SE, Bakken BH, Blackman BK, Chapman MA, Burke JM (2012) The sunflower (*Helianthus annuus* L.) genome reflects a recent history of biased accumulation of transposable elements. The Plant J 72(1):142–153

Tagashira N, Kondo K (2001) Chromosome phylogeny of Zamia and Ceratozamia by means of Robertsonian changes detected by fluorescence in situ hybridization (FISH) technique of rDNA. Plant Syst Evol 227(3–4):145–155

Tek AL, Kashihara K, Murata M, Nagaki K et al (2010) Functional centromeres in soybean include two distinct tandem repeats and a retrotransposon. Chrom Res 18(3):337–347

The International Barley Genome Sequencing Consortium (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711–716

Tran TD, Šimková H, Schmidt R, Doležel J, Schubert I et al (2016) Chromosome identification for the carnivorous plant *Genlisea margaretae*. Chromosoma 126(3):389–397

Ugarković D (2009) Centromere-competent DNA: Structure and evolution. Progr Mol Subcell Biol 48:53–76

Wessler SR (1996) Plant retrotransposons: turned on by stress. Curr Biol 6(8):959–961

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Schuiman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8 (12):973–982

Xu Z, Wang H (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res 35:W265–W268

Xun X, Pan S, Cheng S, Bo Z, Visser RGF (2011) Genome sequence and analysis of the tuber crop potato. Nature 475(7355):189–195

Zhao R, Miao H, Song W, Chen C, Zhang H (2018) Identification of sesame (*Sesamum indicum* L.) chromosomes using the BAC-FISH system. Plant Biol 20:85–92

Genome Annotation and Gene Families in Sesame

15

Hongmei Miao, Yamin Sun, Chun Li, Lei Wang, and Haiyang Zhang

Abstract

In genomes, genes are the most important components because of the function in controlling various biological traits and regulating the complex biological processes in plants. Reliable chromosome-scaled genome assembly supplies the abundant and precise genome information for gene and gene family research in sesame. In this chapter, we deliberate on the main techniques used for genome annotation and gene prediction in sesame (var. Yuzhi 11). Distribution of all genes and gene families in sesame genome is described. Meanwhile, the main research achievements in some key gene families related to key biological processes, such as fatty acid biosynthesis and metabolism and responses to environmental conditions in sesame, are discussed in the section.

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

Y. Sun · L. Wang
TEDA School of Biological Sciences and Biotechnology, Nankai University, Tianjin, P. R. China
e-mail: nksunyamin@aliyun.com

L. Wang
e-mail: wanglei@nankai.edu.cn

15.1 Introduction

The development of high-throughput sequencing technologies facilitates the genome sequencing and assembly and finally realizes the annotation of all the genes in genomes. Besides repetitive sequences, genes are the most important components in higher plant genomes. Before analyzing the genome structure and gene families, all the components including repetitive DNA sequences, coding gene sequences, and non-coding RNAs should be precisely annotated in the sesame reference genome.

In general, genome annotation covers the following processes: identification of repeat sequences, prediction of non-coding RNAs, identification and annotation of coding genes, prediction of gene function, and annotation of pseudogenes. Especially, gene function prediction is to find out the association between descriptions of the biological function with an annotated gene. Since the whole genome sequence produces huge amount of data, it is difficult or impossible to test the function of all the genes. Therefore, alignment of homologous genes is usually used to predict gene function. Some gene datasets, such as NCBI, KEGG, Uniprot, GeneOntology, and EMBL, are commonly used to improve the gene prediction quality.

In the last chapter, the major types and distribution characters of repetitive sequences have

H. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-98098-0_15

been discussed. Homolog-based and ab initio methods were applied to identify the interspersed repeats. RepeatMasker (Maja and Nansheng 2009) and RepeatModeler (Grundmann et al. 2009) (http://www.repeatmasker.org/RepeatModeler.html) are mainly used for the analysis. Tandem Repeats Finder (TRF) (Benson 1999) has been used to screen tandem repeats. After repeat masked, the coding gene sequences and non-coding RNAs in genomes could be annotated. For sesame genome, three approaches, i.e., the homolog-based prediction, de novo prediction, and transcriptome-based prediction, have been used to predict protein-coding genes. All the approaches are described in this chapter. Furthermore, considering sesame is an oilseed crop, some key gene families related to important biological processes are also described in this chapter. Meanwhile, the characters of some key gene families are also described to reflect the species-specific evolution of the *Sesamum* species.

15.2 Genome Annotation Approaches and Gene Prediction

15.2.1 Genome Annotation Approaches

Genome-wide annotation of genes in a species comprises two steps. The first is aligning expressed sequence tags (ESTs), proteins, and other DNA sequences to the genome and ab initio and/or obtaining the evidence-driven gene predictions. The other is synthesizing the gene prediction data into gene annotation dataset (Yandell and Ence 2012). During genome annotation, repeat sequences should be identified at the same time. Thus, some tool and pipelines such as Ensembl can perform the annotation of protein-coding genes and also be used for annotating non-coding RNAs (ncRNAs) (Cunningham et al. 2015).

There are five basic categories of gene annotation software, i.e., (1) Ab initio and evidence-drivable gene predictors, such as Augustus (Stanke and Waack 2003; Stanke et al. 2006a, b) and SNAP (Korf 2004); (2) aligners and assemblers for EST, protein, and RNA-seq, such as BLAST (Altschul et al. 1990, 1997; Korf et al. 2003) and Trinity (Grabherr et al. 2011); (3) Choosers and combiners, such as JIGWAW (Allen and Salzberg 2005) and EVidenceModeler (Haas et al. 2008); (4) Genome annotation pipelines, such as PASA (Haas et al. 2003, 2011), NCBI (Kitts 2002), and Ensembl (Curwen et al. 2004); and (5) Genome browsers for curation, such as Artemis (Rutherford et al. 2000) and JBROWSE (Skinner et al. 2009; Yandell and Ence 2012). Of which homolog-based, ab initio, and transcriptome-based predictions are the most popular methods for gene structure prediction (Cao et al. 2019).

Prediction data indicated that the results of gene annotation are affected by genome assembly completeness, annotation software, and pipelines (Yandell and Ence 2012). Large amounts of gene annotation information and homologous genes enhanced the complication of gene prediction. Thus, by the early 2020, the three famous genome databases including DNA Data Bank of Japan (DDBJ)/European Nucleotide Archive (ENA)/National Center for Biotechnology Information (GenBank) requested that all the genomes should be annotated with the same naming conventions and annotation standard before uploaded to the databases for reservation and releasing (https://www.ncbi.nlm.nih.gov/genbank/eukaryotic_genome_submission_annotation.html#CDS).

15.2.2 Non-coding RNA Prediction

Non-coding RNAs (ncRNAs) are RNAs which could not be translated into protein, including a varied of types, such as tRNA, rRNA, snRNA, and miRNA. ncRNAs are of great biological functions from regulating gene expression and vital cellular functions to affect genome structure (Hüttenhofer et al. 2005). As the ncRNA type is various with different characteristics, many ncRNA prediction softwares are developed for individual ncRNA types. For example, tRNAScan-SE (Lowe and Eddy 1997) is

designed for tRNA prediction, while snoScan (Lowe and Eddy 1999) is used to find snoRNAs carrying C/D box. Later, a database Rfam was developed, together with the other ncRNA tools (Gardner et al. 2015). The database contains more than 1600 RNA families and could be applied to accurately identify many ncRNA types.

15.2.3 Coding Gene and Pseudogene

The structure of a gene consists of many elements such as enhancer, promoter, 5'UTR, open reading frame, 3'UTR, terminator, and other regulating elements (Pearson 2006). Performing coding gene prediction is time consuming during genome annotation. After ab initio, protein homolog-based, and cDNA/transcriptome-based predictions, all the information is harvested and merged together to obtain a final set of gene annotations (Brandi et al. 2008; Haas et al. 2008). In the gene annotation, some pseudogenes could be found. Pseudogenes are segments of DNA that are related to real genes. Compared with the normal genes, pseudogenes have some gene segments but lost all or some functionality, due to the accumulation of multiple mutations or gene copy variation (Vanin 1985). Thus, pseudogenes possess some regulatory functions, but are not the necessary genes required for the survival of an organism.

15.3 Sesame Genome Annotation

In the Sesame Genome Project, the first reference genome assembly (var. Yuzhi 11) was reported 293.7 Mb in length, with a GC content of 34.65%. The N50 and N90 sizes of the scaffolds are 22.6 kb and 4.3 kb, respectively (Zhang et al. 2013). Gene prediction was performed using InchWorm (Grabherr et al. 2011). About 3.5 Gb of RNA-Seq reads (NCBI Accession: SRX061117) were translated into 82,549 peptides for gene prediction (E-value: 1e−5) against the SWISS-PROT (Boeckmann et al. 2003). The GMAP mapping results were applied for ab initio prediction using Augustus (Stanke et al. 2006a, b). As a result, a total of 23,713 gene models were obtained with a total length of 28 Mb. Average gene length was 1.2 kb, and average GC content was 45% (Zhang et al. 2013).

During the chromosome-scale genome annotation (version 3.0), three approaches including the homolog-based, ab initio, and transcriptome-based predictions were used to predict protein-coding genes in sesame genomes (data not shown). For homolog-based prediction, before annotating the sesame genome, the known genomes of five plants (i.e., *Arabidopsis thaliana*, *Glycine max*, *Vitis vinifera*, *Ricinus communis*, and *Solanum lycopersicum*) and homologous proteins were downloaded from JGI (Joint Genome Institute of the US Department of Energy (DOE)) (https://jgi.doe.gov/) and applied for alignment to the repeat-masked *Sesamum* genomes using tblastn (version 2.2.26+) (Kent 2002). Based on the homology alignments of proteins to the genome, the gene structure could be generated using Genewise (version 2.2.0) (Birney et al. 2004).

For ab initio gene prediction, the repeat-masked genome sequences were applied as inputs using Augustus (version 2.5.5) (Stanke and Waack 2003), Genescan (version 1.0) (Aggarwal and Ramaswamy 2002), GlimmerHMM (version 3.0.1) (Majoros et al. 2004), and SNAP15 (Korf 2004). For RNA-seq-based gene prediction, samples of sesame were collected for RNA sequencing using Illumina method. A total of 10.26 Gb RNA-seq data were mapped to the genome using Tophat (version 2.0.8) (Kim et al. 2013. Cufflinks (version 2.1.1) (http://cufflinks.cbcb.umd.edu/) was used to identify spliced transcripts to the gene models.

Subsequently, all gene evidence predicted from the above three approaches were combined into a weighted and non-redundant consensus of the gene structures using EVidenceModeler (EVM) (Haas et al. 2008). Gene models generated by EVM were filtered according to the following criteria: coding region lengths of ≤ 150 bp, supported only by ab initio methods and with FPKM <5, hit with uniref90

database. Gene structure was analyzed using Genewise (version 2.2.0) (Birney et al. 2004).

Meanwhile, BLASTp was used to align the predicted protein-coding genes in sesame to NCBI-nr (of non-redundant protein sequences) (Benson et al. 2005), SwissProt databases (http://www.gpmaw.com/html/swiss-prot.html), and Pfam (Bateman 2004) database with an e-value of 1e−5. The criteria of the identity ≥ 0.25 and the minimum alignment length of 100 bp were set. InterProscan (Jones 2004) was introduced to annotate motifs and domains in gene sequences through comparisons with publicly available databases. Putative gene pathways were derived from genes matched in the KEGG database. The Gene Ontology (GO) (Ashburner 2000) information for each gene code was extracted from InterProscan results. Putative gene pathways were derived from genes matched in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al. 2004).

In the chromosome-scaled genome assembly (version 3), a total of 31,462 genes were detected and annotated. The average gene length was 2620 bp. The gene sequences occupied 24.6% of the genome size (unpublished data, Haiyang Zhang).

15.4 Genome Completeness Assessment

During assessing the quality of a genome assembly, so many indicators such as the base content, scaffold number and length (including N50 and N90), short-inset read mapping rate, short read coverage, single nucleotide polymorphisms (SNP) ratio are used to validate the comparability, uniformity, reliability, and accuracy of the genome. As to assessing the completeness of the genome assembly, Core Eukaryotic Genes Mapping Approach (CEGMA) (http://korflab.ucdavis.edu/dataseda/cegma/) and Benchmarking Universal Single-Copy Orthologs (BUSCO) (http://busco.ezlab.org/) are often used.

For sesame, the Sesame Genome Working Group performed gene annotation validation using BUSCO (Waterhouse et al. 2017), and the core genes were also assessed using CEGMA (Parra et al. 2007) at the same time (Unpublished data, Haiyang Zhang). 97.98% core orthologs were completely detected in sesame reference genome. Meanwhile, of the 248 conserved core eukaryotic genes, more than 93.55% genes were completely covered in reference genome. Both BUSCO and CEGMA assessment indicated that the reference genome was highly complete in terms of protein-coding sequence.

15.5 Gene Distribution and Function in Sesame Genome

15.5.1 Gene Distribution in Sesame Genome

In order to clarify the distribution of genes, the SGWG calculated the gene density per chromosome and analyzed the distribution characters of the sesame genes (Table 15.1) (Unpublished data, Haiyang Zhang). Of the 31,462 genes in sesame genome, *Si*Chr.1 carries the highest gene number (2922) with the high gene sequence percentage (26.86%), while *Si*Chr.8 carries 2911 genes with the highest gene sequence percentage (40.74%). *Si*Chr.12 contains 1693 genes with the gene sequence percentage of 17.6%. The lowest gene number might be related with the short chromosome length.

Furthermore, the gene density was estimated on the basis of the number of genes in non-overlapping 200 kb windows (Fig. 15.1). The gene density in sesame ranged from 0 to 45 genes per 200 kb. The uneven distribution character of sesame genes on chromosomes reflects the genome structure. Comparison results showed that most of the low or zero gene density regions overlapped with centromere positions (data not shown). In order to explore the genome

Table 15.1 Distribution of the predicted genes on 13 chromosomes of *S. indicum*

Chromosome no	Gene number	Total gene length (bp)	Gene sequence percentage (%)
*Si*Chr.1	2922	7,711,294	26.86
*Si*Chr.2	1783	4,681,714	15.97
*Si*Chr.3	1731	4,945,195	19.95
*Si*Chr.4	2055	4,701,923	15.93
*Si*Chr.5	2104	5,388,244	21.37
*Si*Chr.6	2638	7,954,015	40.72
*Si*Chr.7	2594	7,314,362	30.66
*Si*Chr.8	2911	8,116,303	40.74
*Si*Chr.9	2494	6,803,584	29.84
*Si*Chr.10	1984	5,681,264	32.50
*Si*Chr.11	2160	5,820,193	31.32
*Si*Chr.12	1693	4,136,525	17.60
*Si*Chr.13	2097	5,231,983	25.63
Other scaffolds	2296	3,938,669	12.48
Total	31,462	82,425,268	24.59

Note Other scaffolds refer to the scaffolds unanchored onto the 13 chromosomes. All the data are supplied by Haiyang Zhang

structure and chromosome karyotype evolution in sesame, genome synteny analysis should be performed.

15.5.2 Gene Function in Sesame Genome

Based on the GO database (Gene Ontology Consortium, 2004) (http://www.geneontology.org/), all the 31,462 genes were classified into three main categories (i.e., 'molecular function', 'biological process', and 'cellular component' categories) (Fig. 15.2). A total of 28,415 (41.53%) genes were classified into the category 'biological process' of which 8396 (21.86%) genes belonged to the 'metabolic process' group, followed by the 'cellular process' (8261, 21.50%). In the second category of 'molecular function', 21,761 (31.81%) genes were further classified into the 15 groups. The top group was 'binding' (10,484, 15.32%). Meanwhile, the third category of 'biological process' contained 18,239 (26.66%) genes of which the 'cell' group ranked the first with 3523 (5.15%) genes.

To understand the species evolution character, the SGWG screened the orthologous genes and paralogous gene clustering between sesame and the 15 crops from the gene annotation data and performed the phylogenetic position analysis based on the chromosome-scaled sesame assembly of sesame (Unpublished data, Haiyang Zhang). Some key results are introduced in the following chapter.

15.6 Gene Family Analysis of Sesame

15.6.1 Gene Family Comparison of Sesame with Other Plants

In the late stage of the Sesame Genome Project, the SGWG applied the assembled genome (version 3.0, unpublished data, Haiyang Zhang) in genome comparison to learn the genome models of sesame. Five genomes of plants, i.e., *R. communis*, *G. max*, *A. thaliana*, *V. vinfera*, and *S. lycopersicum* were downloaded in JGI (https://

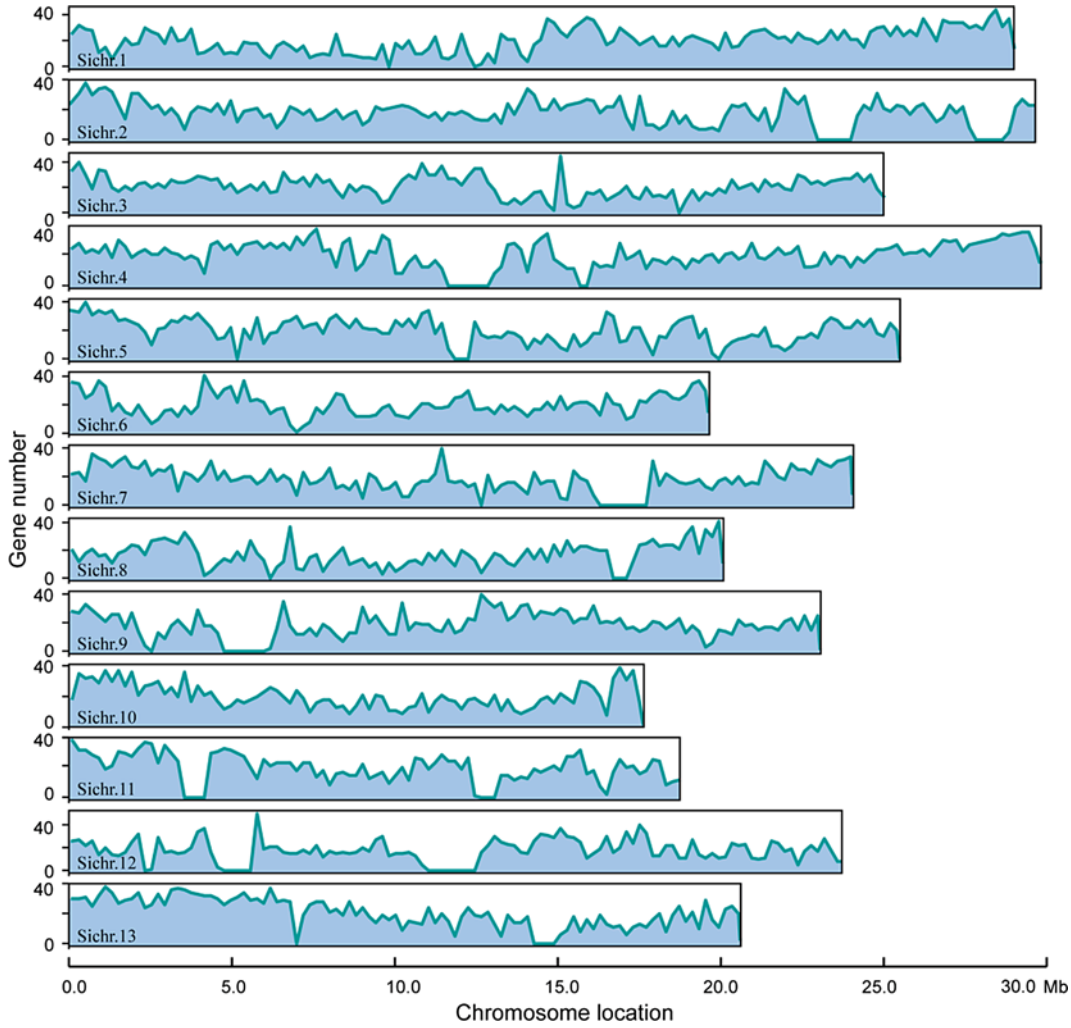


Fig. 15.1 Gene distribution of sesame genome. Green block indicates the gene density of each chromosome, which is calculated according to the gene number per window size of 200 kb

jgi.doe.gov/). In order to perform proper alignment, poor-quality sequences models were filtered with the following cutoff: shorter than 30 amino acids and possessing stop codon percentage above 20% (Table 15.2). The filtering result was shown in Table 15.2.

Average identity matrix was generated using all the homologous genes of the six species and Get_Homologues (Vinuesa and Contreras-Moreira 2015). The homologous genes relationship analysis between sesame and the other five species was performed using two methods, i.e., Get_Homologues named COGS (Kristensen et al. 2010) and OMCL (Li et al. 2003). The results showed that the genome similarity of *S. indicum* and other species is similar. Two group results showed that 90,393 and 98,867 clusters were obtained though COGS and OMCL, respectively (Fig. 15.3). The core gene families (shared by all the six species) and unique clusters in sesame were also identified (Table 15.3) (Unpublished data, Haiyang Zhang).

Enrichment of the 5,665 unique clusters shown in COGS and OMCL was performed in terms of Pfam domains. The results indicated that only three entries were significant enriched at the

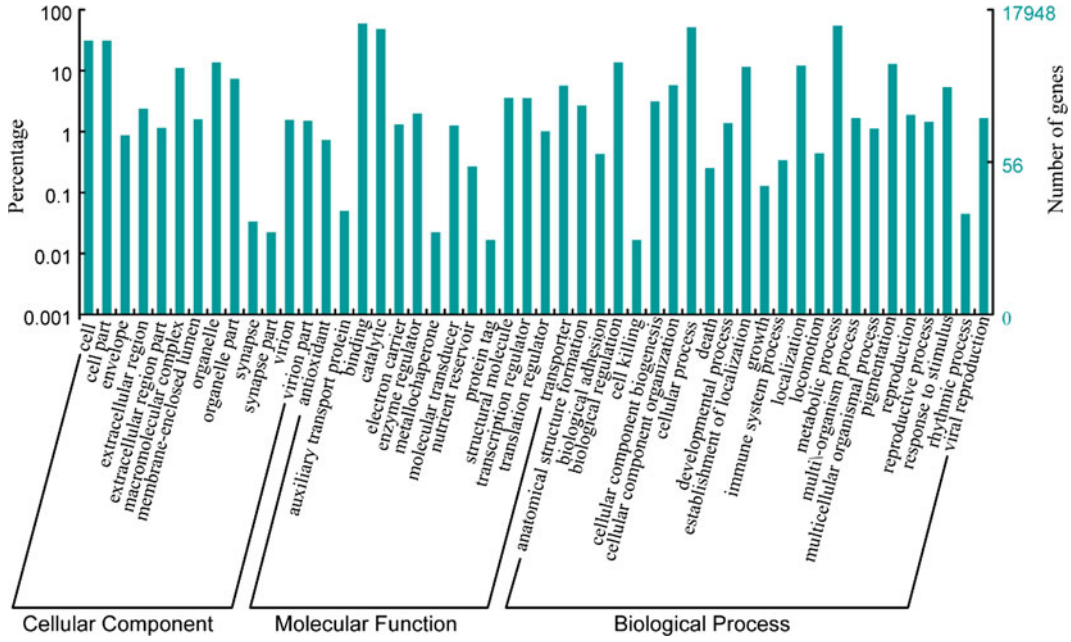


Fig. 15.2 GO classification of the sesame genes. The left vertical axis indicates the gene percentage. The right vertical axis indicates gene number. The horizontal axis indicates gene group type and category (Provided by Haiyang Zhang)

Table 15.2 Statistics of the filtered gene models for genome comparison

Species	Annotation version	Gene number	Filtered data
A. thaliana	TAIR10	35,386	35,350
R. communis	V0.1	31,221	31,221
G. max	Wm82.a2.v1	88,647	88,646
V. vinifera	12X	26,346	26,129
S. lycopersicum	iTAGv2.3	34,727	34,442
S. indicum (var. Yuzhi 11)	Version 3.0	31,462	31,335

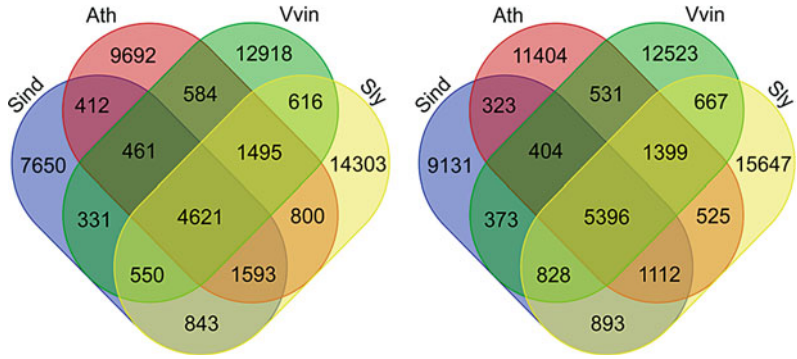


Fig. 15.3 Venn diagram of the homologous clusters between sesame and the three plants. Left figure is determined using COGS method; right figure is determined by OMCL method. Only four species are listed and shown in the figure

Table 15.3 Homologous gene clusters of sesame and the five species

Method	Total clusters	Core cluster	Unique cluster in sesame
COGS	90,393	3985	7023
OMCL	98,867	4560	8659

Table 15.4 Enriched pfam in sesame unique genes

Pfam ID	5665 unique cluster count	Counts in control	p-value (FDR-adjusted)	description
PF01453	5	204	1.00E−01	D-mannose binding lectin
PF00954	3	148	2.74E−01	S-locus glycoprotein
PF08276	3	146	2.74E−01	PAN-like domain

level of 0.5 (FDR-adjusted p-value) (Table 15.4), which suggested that the sesame unique genes are from varied type of function.

Recently, based on the chromosome-scaled genome assembly (version 3.0) and other 21 species, the SGWG constructed 38,382 gene clusters (BlastP, P <E−5) (Unpublished data, Haiyang Zhang). A total of 1172 gene families expanded and 1549 families contracted in sesame (data not shown). Main gene families were involved in zeatin biosynthesis, second metabolites biosynthesis, ribosome, phytosynthesis, oxidative phosphorylation, signaling pathway, plant–pathogen interaction, phynylpropanoid biosynthesis, cysteine and methionine metabolism, and other important pathways. Based on KEGG enrichment analysis, the significantly expanded and contracted gene families (P <0.05) were collected using OmicShare Tools (https://omictools.com/) for further key gene family analysis. All the analysis results supply precise information for further genomics and genetics research in sesame.

15.6.2 FA Gene Families in Sesame

Compared with oilseed rape (*Brassica napus*), soybean (*Glycine max*), peanut (*Arachis hypogaea*), sunflower (*Helianthus annuus*), and other oilseed crops, sesame seed contains high oil content. To reveal the genetic characters of oil biosynthesis, the WGSG screened all the genes related with fatty acid synthesis and metabolism (abbr. FA genes) and constructed FA gene database of *Sesamum* species (data not shown). During FA gene family analysis, the acyl lipid metabolism reference dataset was downloaded from ARALIP (http://aralip.plantbiology.msu.edu/pathways/pathways) and applied for FA gene family detection in *Sesamum* species. BlastPprogram was used to align all candidate FA genes in sesame to the ARALIP database (p <e−5, protein identity >60%). Five oil seed crops, i.e., *H. annuus*, *G. max*, *G. raimondii*, *L. usitatissimum*, *R. communis* were applied for FA gene homolog comparison. There are 677 FA genes in sesame genome. The ratio of FA gene in genomes is 2.15%, similar to those (1.92–2.82%) of above oilseed crops (Unpublished data, Haiyang Zhang).

Gene family comparison results indicated that five FA gene families were significantly expanded, and seven were significantly contracted in *Sesamum*. However, Ka/Ks comparison results showed that no FA gene families in sesame received positive nature pressure (Ka/Ks >1). The results suggested that the high oil content of sesame seeds results from the variation of genetics basis of individual species rather than environment change (data not shown). According to the timeline of the Sesame Genome Project, the SGWG group will publish all the results soon.

15.6.3 R Gene Families in Sesame

Sesame has high tolerance to drought and high temperatures, but the resistance to several key diseases such as *Fusarium* wilt and charcoal rot diseases is relatively low. Except the wild species, no cultivars are immune to pathogens of *Fusarium* wilt and charcoal rot diseases (Joshi 1961; Uzo 1985; Liu et al. 1993; Nimmakayala et al. 2011; Qiu et al. 2014; Zhang et al. 2019). Identification of R genes and their pathogen effectors is essential for understanding host–pathogen interactions and improving disease resistance breeding in sesame. Based on the plant resistance gene (PRG) database (PRGdb 3.0) (Osuna-Cruz et al. 2017), the SGWG systematically screened the R gene homologs in the sesame assembled genome constructed the disease resistance genes (R genes) data was constructed for sesame (Unpublished data, Haiyang Zhang). There are 177,072 putative PRGs and 153 reference PRGs and 23 avirulence genes in PRGdb. In sesame genome, 1198 R genes (*p*-value <1e−06) were detected and grouped into 13 R gene types of which KIN (Kinase), RLK (Receptor like Kinase), RLP (Resistance to powdery mildew), and NBS (nucleotide-binding subdomain) type are the top R gene types (Unpublished data, Haiyang Zhang). Expansion and contraction analysis of gene families indicated that five key R gene families were significantly contracted in sesame which might result in the low resistance to fungi pathogens (data not shown). The results suggest that enhancing the function of above key orthologous R genes should increase the resistance to disease pathogens in sesame.

15.6.4 Other Key Gene Families in Sesame

In recent a few years, based on sesame genome, several transcription factor (TF) families, such as Bzip transcription factor, MYB, and WRKY families, related to important biological processes were analyzed (Wei et al. 2015; Dossa et al. 2016a, b; Chowdhury et al. 2017; Li et al. 2017; Mmadi et al. 2017; Wang et al. 2018; Wei et al. 2019). For example, Wei et al. (2015) analyzed the transcription factor MADS-box gene family in sesame for the first time. MADS represents the four homologs in different organisms, i.e., MCM1 (Mini Chromosome Maintenance 1) of yeast (*Saccharomyces cerevisiae*), AG (Agamous) of *Arabidopsis thaliana*, DEF (Deficiens) of snapdragon (*Antirrhinum majus* L.), and SRF (Serum Response Factor) of humans (Wei et al. 2015). In the de novo sesame genome assembly (containing 24,148 genes in 14 linkage groups) (Wang et al. 2014), 57 non-redundant MADS-box proteins were identified using both SRF (type I) and MEF2 (type II) MADS-box domain sequences as queries. All the 57 MADS-box genes were grouped into four groups according to the phylogenetic tree of sesame and other three plants. Expression profile indicated that MIKCC-type MADS-box genes might play roles in sesame flower and seed development.

As to the one of the largest transcription factors (TFs), MYB gene family, Li et al. (2017) screened 287 MYB genes and found some *SIMYBs* regulated the responses to drought and waterlogging in sesame. Wei et al. (2019) analyzed the 45 homeodomain-leucine zipper (HD-Zip) gene family in sesame and detected 75% HD-Zip genes presented the responses to drought and salinity treatment.

Till now, tens of gene families have been analyzed in sesame genome (Wei et al. 2017; You et al. 2018; Miao et al. 2019; Yu et al. 2019; Zhang et al. 2019). The above research results reflect the function of transcription factors in sesame. Recently, in order to promote the functional genomics research of sesame, Wei et al. (2017) combined genetic and comprehensive phenotypic information and constructed the Web-based database Sesame FG. A large amount of genome information and genome re-sequencing data supply the solid foundation for genomics and genetic analysis in sesame.

15.6.5 Gene Families in Pan Genomes of Sesame

Aside from the Sesame Genome Project and the chromosome-scaled genome assembly (var. Yuzhi 11) (Zhang et al. 2013; partially unpublished data, Haiyang Zhang), Wang et al. (2014) constructed a draft genome with 16 linkage groups (var. Zhongzhi no. 13) (Deposited no. SRA122008) using Illumina sequencing platform (Wang et al. 2014, 2016). The draft genome size was 274 Mb containing 24,148 genes encoding 27,148 predicted proteins in 16 linkage groups. Subsequently, Wei et al. (2015) performed draft genome assemblies for two sesame landraces (i.e., Baizhima and Mishuozhima). The National Bureau of Plant Genetic Resources of India assembled a genome draft (cv. Swetha) to 340 Mb (Kitts et al. 2016).

To realize gene discovery and improve the genome information, Yu et al. (2019) constructed a sesame pan-genome of 554.05 Mb with the above five genome datasets using the reference-assisted assembly approach. The pan-genome contained 26,472 orthologous gene clusters and 15,409 (58.21%) core clusters (common in all five sesame genomes). The variation of gene families and gene numbers between the five genomes reflected the geography difference of materials, genome sequencing data, and the genome variation. Genome evolution analysis inferred that some genes involved in plant–pathogen interaction and lipid metabolism bore positive selection. The genomics studies supply more information for sesame molecular breeding.

References

Aggarwal G, Ramaswamy R (2002) *Ab initio* gene identification: prokaryote genome annotation with GeneScan and GLIMMER. J Biosci 27(1):7–14

Allen JE, Salzberg SL (2005) JIGSAW: integration of multiple sources of evidence for gene prediction. Bioinformatics 21:3596–3603

Altschul SF, Madden TL, Alejandro S, Zhang J, Zhang Z et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Ashburner M, Ball CA, Blake JA, Botstein BH et al (2000) Gene ontology: tool for the unification of biology. Nat Genet 25:25–29

Benson D, Karsch-Mizrachi I, Lipman D, Ostell J, Wheeler D (2005) GenBank. Nucleic Acids Res 1 (33 (Database issue)):D34–38

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27:573–580

Birney E, Clamp M, Durbin R (2004) Gene wise and genomewise. Genome Res 14(5):988–995

Brandi LC, Ian K, Sofia MC, Genis P, Eric R et al (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18(1):188–196

Chowdhury S, Basu A, Kundu S (2017) Overexpression of a new osmotin-like protein gene (SindOLP) confers tolerance against biotic and abiotic stresses in sesame. Front Plant Sci 8:410

Cunningham F, Amode MR, Barrell D, Beal K, Billis K et al (2015) Ensembl 2015. Nucleic Acids Res 43(D1): D662–D669

Curwen V, Eyras E, Andrews TD, Clarke L, Mongin E et al (2004) The Ensembl automatic gene annotation system. Genome Res 14:942–950

Cao D, Wang M, Ge1 Y, Gong S (2019) Draft genome of the big-headed turtle *Platysternon megacephalum*. Sci Data 6(1):60

Dossa K, Diouf D, Cissé N (2016) Genome-wide investigation of Hsf genes in sesame reveals their segmental duplication expansion and their active role in drought stress response. Front Plant Sci 7:1522

Dossa K, Wei X, Li D, Fonceka D, Zhang Y et al (2016) Insight into the ap2/erf transcription factor superfamily in sesame and expression profiling of DREB subfamily under drought stress. BMC Plant Biol 16 (1):171

Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL et al (2015) Rfam: updates to the RNA families database. Nucleic Acids Res 37(Database issue): D136–40

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652

Grundmann N, Demester L, Makalowski W (2009) TEclass—a tool for automated classification of unknown eukaryotic transposable elements. Bioinformatics 25(10):1329–1330

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE et al (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble apliced alignments. Genome Biol 9(1):R7

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr et al (2003) Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res 31:5654–5666

Hüttenhofer A, Schattner P, Polacek N (2005) Non-coding RNAs: hope or hype? Trends Genet 21 (5):289–297

Joshi AB (1961) Sesamum. Indian Central Oilseed Committee Hyderabad, India, p 109

Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. Nucleic Acids Res 32(Database issue): D277-D80

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R et al (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14:R36

Kitts P (2002) Genome assembly and annotation process. In: McEntyre J, Ostel (eds) The NCBI handbook, National Center for Biotechnology Information, Chap. 13

Kitts PA, Church DM, Françoise TN, Jinna C, Vichet H et al (2016) Assembly: a resource for assembled genomes at NCBI. Nucleic Acids Res D1:D73–D80

Korf I (2004) Gene finding in novel genomes. BMC Bioinformatics 5(1):59

Korf I, Yandell M, Bedell J (2003) BLAST: an essential guide to the basic local alignment search tool, vol 339. O'Reilly & Associates

Kristensen DM, Kannan L, Coleman MK, Wolf YI, Sorokin A et al (2010) A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. Bioinformatics 26(12):1481–1487

Li D, Liu P, Yu J, Wang L, Dossa K et al (2017) Genome-wide analysis of WRKY gene family in the sesame genome and identification of the WRKY genes involved in responses to abiotic stresses. BMC Plant Biol 17(1):152

Li L, Stoeckert CJ Jr, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13(9):2178–2189

Liu J, Tu L, Xu R, Zheng Y (1993) The relationship between the waterlogging resistance and the genotypes and the vigor of root system in sesame (*Sesamum indicum* L.). Acta Agri Bor-Sin 8(3):82–86

Lowe TM, Eddy SR (1997) tRNAScan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964

Lowe TM, Eddy SR (1999) A Computational screen for methylation guide snoRNAs in yeast. Science 283 (5405):1168–1171

Maja TG, Nansheng C (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinform 25, Chap. 4, Unit 4.10. https://doi.org/10.1002/0471250953.bi0410s25.

Majoros WH, Pertea M, Salzberg SL (2004) TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. Bioinformatics 20(16):2878–2879

Miao H, Li C, Duan Y, Wei L, Ju M et al (2019) Identification of a *Sidwf1* gene controlling short internode length trait in the sesame dwarf mutant *dw607*. Theor Appl Genet 133(1):73–86

Mmadi MA, Dossa K, Wang L, Zhou R, Wang Y et al (2017) Functional characterization of the versatile MYB gene family uncovered their important roles in plant development and responses to drought and waterlogging in sesame. Genes 8(12):362

Nimmakayala P, Perumal R, Mulpuri S, Reddy UK (2011) *Sesamum*. In: Kole C (ed) Wild crop relatives: genomic and breeding resources. Volume oilseeds. Springer, Berlin Heidelberg, pp 261–273

Pearson H (2006) Genetics: what is a gene? Nature 441 (7092):398–401

Parra G, Bradnam K, Korf I (2007) CEGMA. Bioinformatics 23:9

Qiu C, Zhang H, Chang S, Wei L, Miao H (2014) Laboratory detecting method for pathogenicity of *Fusarium oxysporum* Schl. f. sp. *sesami* isolates. Acta Phytopathol Sin 44(1):26–35

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P et al (2000) Artemis: sequence visualization and annotation. Bioinformatics 16:944–945

Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH (2009) JBrowse: a next generation genome browser. Genome Res 19:1630–1638

Stanke M, Keller O, Gunduz I, Hayes A, Waack S et al (2006) AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res 34:W435–W239

Stanke M, Waack S (2003) Gene prediction with a hidden Markov model and a new intron submodel. Bioinform Suppl 2:ii215–225

Stanke M, Schoffmann O, Morgenstern B, Waack S (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinform 7:62

Uzo JO (1985) A search for drought resistance in the wild relatives of the cultivated sesame (*Sesamum indicum*). In: Ashri A (ed) Sesame and safflower: status and potential. FAO Plant production and protection paper 66, Rome, Italy, pp 163–165

Vanin EF (1985) Processed pseudogenes: characteristics and evolution. Ann Rev Genet 19:253–272

Wang L, Xia Q, Zhang Y, Zhu X, Zhu X et al (2016) Updated sesame genome assembly and fine mapping of plant height and seed coat color QTLs using a new high-density genetic map. BMC Genom 17:31

Wang L, Yu S, Tong C, Zhao Y, Liu Y et al (2014) Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. Genome Biol 15 (2):R39

Wang Y, Zhang Y, Zhou R, Dossa K, Yu J et al (2018) Identification and characterization of the bZIP transcription factor family and its expression in response to abiotic stresses in sesame. PLoS ONE 13(7): e0200850

Waterhouse RM, Seppey M, Simo FA, Manni M, Zdobnov EM (2017) Busco applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol 35(3):543–548

Wei M, Liu A, Zhang Y, Zhou Y, Li D et al (2019) Genome-wide characterization and expression

analysis of the HD-Zip gene family in response to drought and salinity stresses in sesame. BMC Genom 20(1):74

Wei X, Gong H, Yu J, Liu P, Wang L et al (2017) SesameFG: an integrated database for the functional genomics of sesame. Sci Rep 7(1):1–10

Wei X, Wang L, Yu J, Zhang Y, Li D, Zhang X (2015) Genome-wide identification and analysis of the MADS-box gene family in sesame. Gene 569(1):66–76

Yandell M, Ence D (2012) A beginner's guide to eukaryotic genome annotation. Nat Rev Genet 13 (5):329–342

You J, Wang Y, Zhang Y, Dossa K, Li D et al (2018) Genome-wide identification and expression analyses of genes involved in raffinose accumulation in sesame. Sci Rep 8(1):1–11

Yu J, Golicz AA, Lu K, Dossa K, Zhang Y et al (2019) Insight into the evolution and functional characteristics of the pan-genome assembly from sesame landraces and modern cultivars. Plant Biotechnol J 17(5):881–892

Zhang H, Miao H, Wang L, Qu L, Liu H et al (2013) Genome sequencing of the important oilseed crop *Sesamum indicum* L. Genome Biol 14(1):401

Zhang H, Miao H, Ju M (2019) Potential for adaptation to climate change through genomic breeding in sesame. In: Kole C (ed) Genomic designing of climate-smart oilseed crops. Springer, Cham, Switzerland, pp 374–376

16 Genome Synteny Analysis and Phylogenetic Position of *S. indicum*

Hongmei Miao, Yamin Sun, Wenchao Lin, Lei Wang, and Haiyang Zhang

Abstract

To explore the genome evolution features of sesame, the Sesame Genome Working Group (SGWG) performed synteny block analysis between sesame and the model plants. We here present some key results of detection of duplication and syntenic blocks in the sesame genome. Distribution and function analysis of homologous genes in sesame genome is shown at the same time. In addition, the phylogenetic position of sesame among the flowering plants is depicted in this chapter. The new findings reveal the conservative regions in sesame after the early whole-genome duplication event and elucidate the genome evolution features in eudicots.

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

Y. Sun · W. Lin · L. Wang
TEDA School of Biological Sciences and Biotechnology, Nankai University, Tianjin, People's Republic of China
e-mail: nksunyamin@aliyun.com

W. Lin
e-mail: linwenchao@yeah.net

L. Wang
e-mail: wanglei@nankai.edu.cn

16.1 Introduction

Flowering plants are the dominant land plants in the world. Their genomes exhibit the diversification of various genomes and the evolutionary features involve in genome duplication, diploidization, and chromosomal rearrangement (Salse 2012; Murat et al. 2017). Recently, the most recent common ancestor (MRCA) of modern monocots and eudicots has been reconstructed to decipher the conserved compartments in the genomes of the modern engiosperm species (Murat et al. 2017). For the eudicots, highlights of genome evolution focuses on variation of the genome structure and chromosome karyotype from ancestral to modern species through the gammar whole genome duplication event (Jaillon et al. 2007). Comparison of the assembled genome with model genomes always display the structural conservation and provides insights into the organization, regulation, and evolution of a genome in a genus or family (Salse 2012; Murat et al. 2017).

As described in the former chapters, the orthologous genes (i.e., conserved genes) in genomes are descended from the paleogenomes, while paralogous genes (i.e., duplicated genes) probably originate through diverse rounds of genome or segmental duplication events (Raes et al. 2003; Salse 2012; Jiao et al. 2014). In some angiosperms with few rounds of whole-genome duplication and diploidization, the synteny

H. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-98098-0_16

conservation regions always cover hundreds to thousands of orthologous genes (Paterson et al. 2009). In order to clarify the genome evolution of the cultivated sesame in *Sesamum*, the Sesame Genome Working Group (SGWG) screened the orthologous and paralogous genes in sesame genome and compared the distribution character of the paralogous genes in sesame with model plants. We here introduce the main synteny analysis results of sesame genome with model plant grapevine and present the specific information and genome evolution characters in *Sesamum*.

16.2 Duplication Block Distribution in Sesame Genome

In plant genomes, the conservative synteny regions can be identified through clustering the collinear matched gene pairs. The collinear genes with the specific gene order show the genome conservation (Myburg et al. 2014). In the Sesame Genome Project (SGP), the Sesame Genome Working Group (SGWG) constructed chromosome-scaled genome assembly for sesame through the complicated hybrid assembly strategy, which provides the necessary and precise genome information for further genome evolution analysis in sesame. With the aid fine genome information (335 Mb), the SGWG initially screened the duplication blocks and paralogous genes in sesame genome to explore the genome duplication (E-value <1e-20) (Fig. 16.1) (Unpublished data, Haiyang Zhang).

Based on the high assembled genome information, a total of 3007 genes were detected duplicated in 55 duplication blocks which covered the 52.04% of the sesame genome (Fig. 16.2) (Unpublished data, Haiyang Zhang). In order to explore the paralogous genes in sesame, the SGWG further performed the GO function analysis of the 3007 duplicated genes (Fig. 16.3) (Unpublished data, Haiyang Zhang). The results indicated that 43% of the 3007 duplicated genes in sesame were classified into the category 'Biological process' (Fig. 16.3). Of the 46 groups, 'Binding' was the largest group containing 1282 duplicated genes, followed were 1030 genes in 'Catalytic activity' group of

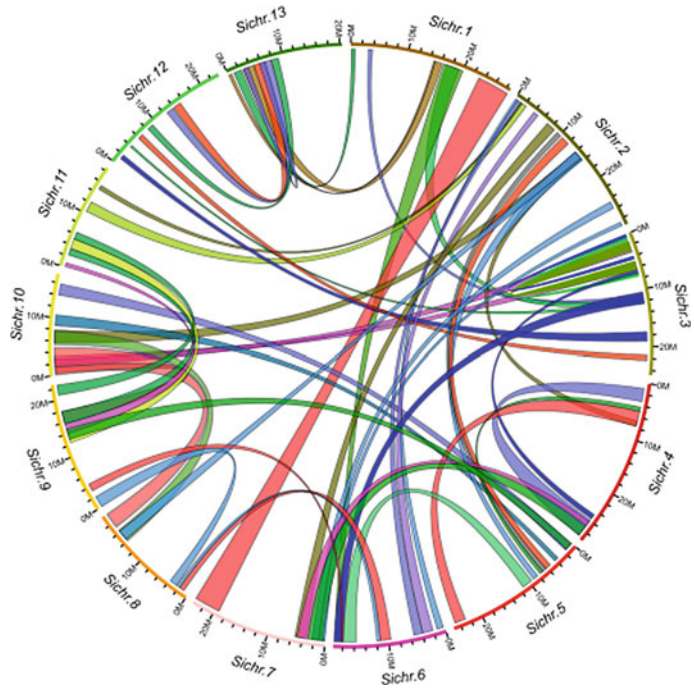


Fig. 16.1 Genome landscape of *S. indicum.* The colored circle refers to the 13 genome sequence molecules which correspond to the 13 chromosomes in sesame. The colored linkage curves indicate the duplication blocks in sesame genome. (Provided by Haiyang Zhang)

category 'Molecular function', and 904 genes in 'Metabolic process' group of category 'Cellular component' and the other groups.

16.3 Genome Synteny Between Sesame and Grapevine

To determine the evolution character of sesame genome, the SGWG compared all the 29,166 sesame genes on the 13 chromosomes (335.19 Mb) with the 24,283 grapevine (*Vitis vinifera*) genes on the 19 chromosomes (427.19 Mb, GCF_000003745.3) (Jaillon et al. 2007) using MCSCANX software with the following standard parameters: MATCH_SCORE: 50; MATCH_SIZE: 20; GAP_PENALTY: −1; OVERLAP_WINDOW: 5; E_VALUE: 1e-05; and MAX GAPS: 25). As a result, a total of 163 synteny blocks containing 14,241 (25.84%) collinear genes of the total 53,449 genes were identified between *S. indicum* and *V. vinifera* (Table 16.1). Of the 14,877 collinear genes, 8123 are sesame genes covering 27.85% of the mapped genes in sesame. The results showed the high conservation of ancestral chromosome fractions in *Sesamum* species.

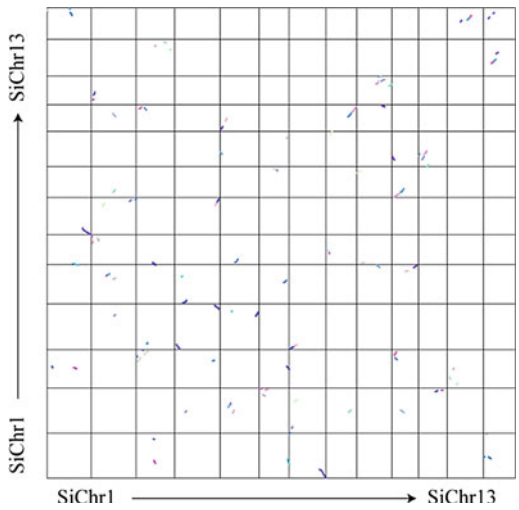


Fig. 16.2 Dotplot diagram of duplication blocks in sesame genome. Horizontal and Vertical blocks indicate the sesame chromosome sequences (*Si*Ch1-*Si*Chi13). (Provided by Haiyang Zhang)

For the 13 sesame chromosomes, the synteny block number per chromosome varied from 5 (in *Si*Ch12) to 19 (in *Si*Ch4) (Table 16.2). The conserved synteny gene number per chromosome ranged from 309 (in *Si*Ch12) to 926 (in *Si*Ch4). Moreover, chromosome screening analysis indicated that all the 8,123 synteny genes were unevenly dispersed onto the 13 chromosomes (Fig. 16.4).

The dotpot diagram showed the complete observed syteny regions between sesame (13 chromosomes with 29,166 genes) and grapevine (19 chromosomes with 24,283 genes) (Fig. 16.4). Compared with the grapevine genome, the sesame genome might undergo more diverse evolution events (such as whole genome duplication and gene loss) and resulted in the low percentage of the ortholgous genes (Data not shown). Meanwhile, the SGWG analyzed the distribution of synteny genes and blocks between sesame and tomato (*Solanum lycopersicum*) genome (version GCF_000188115.3). The results showed the high percentage of collinear genes of sesame which accords with the close phylogenic relationship between sesame and tomato (Data not shown). Thus, to further clarify the genome evolution of sesame, the SGWG performed the phylogenetic analysis using the precise genome assembly data.

16.4 Phylogenetic Position of Sesame

Based on the Angiosperm Phylogeny Group information, sesame has been located in the asterids clade of the core eudicotyledons of Angiosperm Phylogeny Group 2 (APG 2) (Bremer et al. 2003). Chloroplast genomic data indicated that *Sesamum* (Pedaliaceae) is a sister genus to the *Olea* and *Jasminum* (Oleaceae family) clade and represents the core lineage of the *Lamiales* families (Yi and Kim 2011). To further explore the evolutionary position of the *Sesamum* sesame in angiosperm plants, the SGWG performed the phylogenetic analysis of chloroplast genomes between sesame (NCBI no. KC569603) and other 13 dicot species in 2013

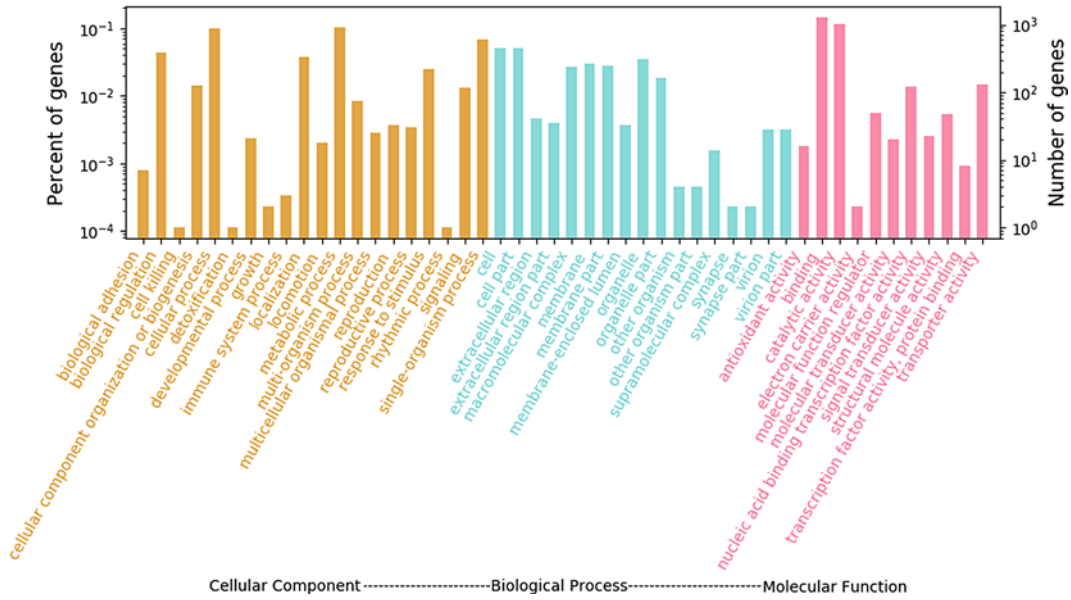


Fig. 16.3 Function identification of the duplication genes in sesame. GO function classification of the 3007 duplication genes in the 55 blocks. All the genes are classified into three categories, i.e., 'Cellular component', 'Biological process', and 'Molecular function'. (Supplied by Haiyang Zhang)

Table 16.1 Genome synteny comparisons of *S. indicum* with grapevine

Species comparison	Chromosome set	Genome size (Mb)	Annotated genes	Genome synteny block statistics			Genome accession
				Synteny gene number	Synteny block number	Percentage of mapped genes (%)	
Vitis vinifera	19	486.20	26,346 (24,283 in 19 chromosomes)	6,754	163 (containing 16,102 genes in blocks)	27.81	GCF_000003745.3
S. indicum	13	335.19	31,462 (29,166 genes in 13 chromosomes)	8,123	163 (containing 14,242 genes in blocks)	27.85	Unpublished version 3.0

*Percentage of genome (%) refers to the ratio of the genes in duplicated blocks to the total genes in 13 chromosomes. (Provided by Haiyang Zhang)

(Zhang et al. 2013a). The *Ka*, *Ks*, and *Ka*/*Ks* ratio of 77 protein-coding genes in sesame and other 13 dicot species from the asterid and rosid clades were compared. The results showed that the 9 genes evolved rapidly, while 3 genes evolved slowly in the sesame chloroplast genome. The conservation of the chloroplast genome suggests a universal evolutionary selection pressure.

Subsequently, the SGWG performed the phylogenic relationship analysis using the genomes of sesame (var. Yuzhi 11, version 1.0) and 36 species from 19 families (Fig. 16.5). The phylogenetic tree showed that *Sesamum* is closely related to the Solanaceae and Phrymaceae families. The above syteny comparison results of sesame and grapevine and tomato genomes prove the phylogenetic position results and reflect the close relationship between *Sesamum* and Sloanaceae species.

At present, the SGWG compared the genomes of *S. indicum* (*cv.* Yuzhi 11) and other 15 crops (i.e., *Aquilegia coerulea*, *Arabidopsis thaliana*,

Table 16.2 Distribution of the synteny blocks between sesame and the grapevine

Chromosome no.	Synteny block number	Synteny gene number
*Si*Chr.1	19	865
*Si*Chr.2	18	848
*Si*Chr.3	18	920
*Si*Chr.4	19	926
*Si*Chr.5	15	669
*Si*Chr.6	10	454
*Si*Chr.7	11	607
*Si*Chr.8	10	464
*Si*Chr.9	11	527
*Si*Chr.10	10	611
*Si*Chr.11	8	396
*Si*Chr.12	5	309
*Si*Chr.13	9	527
Total	163	8,123

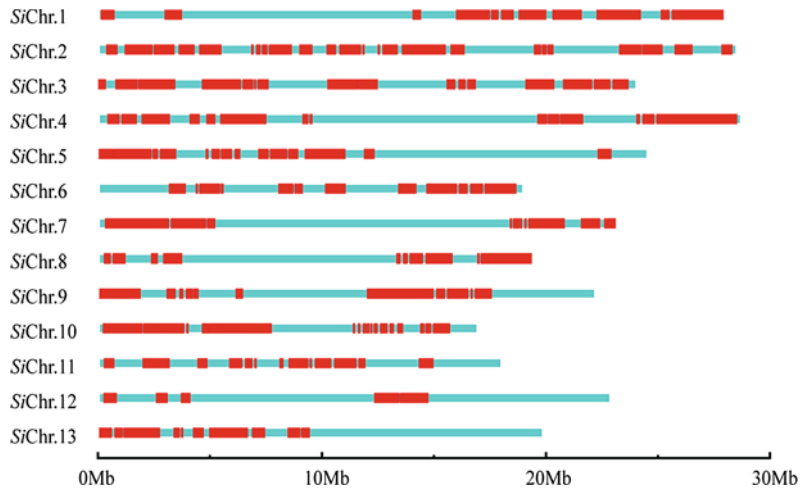


Fig. 16.4 Dotplot diagram of synteny genes between sesame and grapevine. Horizontal blocks indicate the sesame chromosome sequences (*Si*Ch1-*Si*Chi13). The green segment indicates the sesame chromosome sequence (Mb). Red vertical bar refers to synteny genes between sesame and grapevine. (Provided by Haiyang Zhang)

Carica papaya, *Citrus clementina*, *Cucumis sativus*, *Eucalyptus grandis*, *Glycine max*, *Minulus guttatus*, *Oryza sativa*, *Populustrichocarpa*, *Prunus persica*, *Ricinus communis*, *Solanum lycopersicum*, *Theobroma cacao*, and *Vitis viniera*). A total of 231 common single-copy genes were identified in all species and applied to detect the amino acid sequence differences under maximum likelihood through RAxML (Randomized Axelerated Maximum Likelihood) software (https://sco.h-its.org/exelixis/web/software/raxml/index.html) (Stamatakis 2014). As a result, more than twenty thousand amino acid substitutions were found and used to construct the phylogenetic tree (Unpublished data, Haiyang Zhang). The three groups of the key

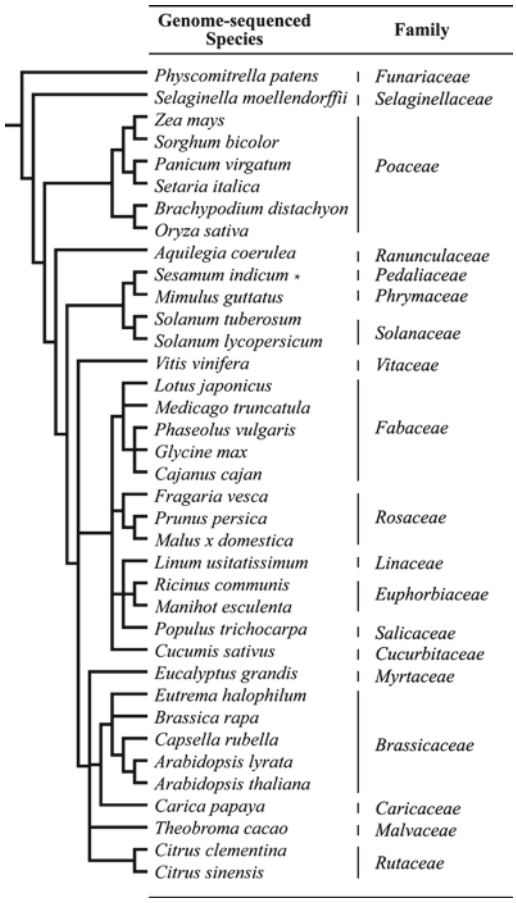


Fig. 16.5 Phylogenetic position comparison of sesame and the 36 plants using genome sequences. *refers to sesame. (Modified from Zhang et al. (2013b))

branch evolution timescales between *Arabidopsis thaliana* and *Solanum lycopersicum* (101.2–123.9 Myr), *Solanum lycopersicum* and *Minulus guttatus* (70–90 Myr), and *Minulus guttatus* and *Vitis viniera* (110–140 Myr) from the NODE TIME dataset (http://www.timetree.org) were applied to determine the divergence time of each species. Finally, the SGWG constructed the phylogenetic tree for sesame and the 16 species (Data not shown). The results showed that the branch with the family Pedaliaceae was divergent at about 121.0 Myr. The family Pedaliaceae (including *Sesamum* and other genera) was divergent at about 47.4 Myr (Unpublished data, Haiyang Zhang).

Moreover, in order to clarify the genome evolution characters of *Sesamum*, the SGWG initiated the genome sequencing and assembly of the wild Sesamum species in 2014. Till now, phylogenetic analysis of the various *Sesamum* species in family Pedaliaceae has been performed based on the fine genomes of the 7 *Sesamum* species. The new findings exhibit that the family Pedaliaceae and the genus *Sesamum* present the important evolutionary position in eudicots (Unpublished data, Haiyang Zhang). To reflect the specific characters of genome evolution in sesame, some information of the wild species is presented in the following chapters. All the new results will be published in near future.

References

Bremer B, Bremer K, Chase MW, Reveal JL, Zmarzty S (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. Bot J Linn Soc 141(4):399–436

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467

Jiao YN, Li JP, Tang HB, Paterson AH (2014) Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. Plant Cell 26(7)(-):2792–2802

Murat F, Armero A, Pont C, Klopp C, Salse J (2017) Reconstructing the genome of the most recent common ancestor of flowering plants. Nat Genet 49(4):490

Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD et al (2014) The genome of *Eucalyptus grandis*. Nature 510(7505):356

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457(7229):551

Raes J, Vandepoele K, Simillion C, Saeys Y, Peer YVD (2003) Investigating ancient duplication events in the Arabidopsis genome. J Struct Funct Genom 3(1–4):117–129

Salse J (2012) In silico archeogenomics unveils modern plant genome organisation, regulation and evolution. Curr Opin Plant Biol 15(2):122–130

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformat 30(9):1312–1313

Yi DK, Kim KJ (2011) Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. PLoS ONE 7(5):e35872

Zhang H, Li C, Miao H, Xiong S (2013a) Insights from the complete chloroplast genome into the evolution of *Sesamum indicum* L. PLoS ONE 8(11):e80508

Zhang H, Miao H, Wang L, Qu L, Liu H et al (2013b) Genome sequencing of the important oilseed crop *Sesamum indicum* L. Genome Biol 14(1):401

Genome Sequencing of the Wild *Sesamum* Speices

17

Haiyang Zhang, Lei Wang, Hongmei Miao, and Yamin Sun

Abstract

The Sesame Genome Working Group sequenced the genomes of the six wild *Sesamum* species, i.e., *S. alatum* (2*n* = 26, var. 3651), *S. latifollum* (2*n* = 32, var. KEN1), *S. angolense* (2*n* = 32, var. K16), *S. calycinum* (2*n* = 32, var. KEN8), *S. angustifolium* (2*n* = 32, var. G01), and *S. radiatum* (2*n* = 64, var. G02), which cover all three types of the basic chromosome number of the genus *Sesamum*. We describe the complicated sequencing and assembly strategies for the wild species and the first interspecific single nucleotide polymorphism (SNP) genetic map for *Sesamum*. The chromosome-scaled genome assemblies of the six wild *Sesamum* species are deliberated in this chapter.

17.1 Introduction

In *Sesamum*, there are 36 species determined according to the index Kewensis (Joshi 1961). Based on chromosome karyotype characters, all the *Sesamum* species are ascribed into three groups, i.e., 2*n* = 26, 32, and 64 (Zhang et al. 2012; Zhao et al. 2018). Some wild species occupy the high economic and academic position in *Sesamum*. For example, *S. angustifolium* (2*n* = 32) has high adaptation for drought resistance, while *S. radiatum* (2*n* = 64) is immune to *Fusarium oxsporum* f. sp. *sesami* and other main diseases and could be used to dig elite disease resistance gene resources (Miao et al. 2019; Zhang et al. 2019). Therefore, the Sesame Genome Working Group intensified the Sesame Genome Project and performed genome sequencing of the six wild *Sesamum* species since 2014 (Table 17.1) (Unpublished data, Haiyang Zhang). We thus summarize the genome sequencing strategies for the wild *Sesamum* species, and introduced the first interspecific genetic maps for *Sesamum*. The basic information of the chromosome-scaled genomes of the six wild species was exhibited accordingly.

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

L. Wang · Y. Sun
TEDA School of Biological Sciences and Biotechnology, Nankai University, Tianjin, People's Republic of China
e-mail: wanglei@nankai.edu.cn

Y. Sun
e-mail: nksunyamin@aliyun.com

H. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-98098-0_17

Table 17.1 Information of the 6 wild *Sesamum* species for genome sequencing

Sample no.	Species name	Sample name	Type
1	*S. indicum* L.	Yuzhi 11	Cultivated sesame, $2n = 26$
2	*S. alatum* Thonn	3651	Wild species, $2n = 26$
3	*S. latifollum* Gillet	KEN1	Wild species, $2n = 32$
4	S. angolense Welw	K16	Wild species, $2n = 32$
5	*S. calycinum* Welw	KEN8	Wild species, $2n = 32$
6	*S. angustifolium* Oliver Engl	G01	Wild species, $2n = 32$
7	*S. radiatum* Schumach. & Thonn. (*S. occidentale*)	G02	Wild species, $2n = 64$

17.2 Genome Sequencing Strategy

Based on the chromosome karyotype and cytogenetic analysis, some wild species were inferred with high genome size due to polyploid character (partial results shown in former chapters). Before initiating the genome sequencing, accurately estimating the genomic characteristics was performed using flow cytometry (BD FACSation™, USA) (Doležel et al. 1989). The genome size of the six wild species ranged from 378 to 723 Mb (Unpublished data, Haiyang Zhang). The Sesame Genome Working Group (WGSG) proposed a complicated large-scale genome sequencing strategies for the six wild species (Fig. 17.1). As shown in Fig. 17.1, Illumina, PacBio, and Hi-C map sequencing platforms were applied. Large amounts of Illumina and PacBio data with more than 30 × coverage sequencing data were generated for each species. In the complicated sequencing and assembly platforms, the de novo assembly of PacBio was combined with Illumina and HiC data to form the high-quality genome assembly using HiC-Pro.

For *S. latifolium* and *S. radiatum*, Bionano mapping was performed to obtain long sequence information. The single-stranded nicking endonuclease *Afl* II was used to prepare the high-molecular-weight DNA. Based on the standard BioNano protocols, the labeled HMW DNA was imaged and scanned on the BioNanoIrys system platform for de novo assembly. Bionano Access software was used to get rid of errors of the physical maps. All the consensus information was applied to reassemble the genome. For *S. calycinum* and *S. angustifolium*, two interspecific single nucleotide polymorphism (SNP) maps were constructed and used to assess the final genome assemblies.

17.3 Interspecific Genetic Map Construction

Studies of closely related species can reveal the genetic basis and maintenance of species divergence (Brennan et al. 2014). In *Sesamum*, incompatibility between the wild species is very common (Joshi 1961; Subramanian 2003; Nimmakayala et al. 2011; Zhang et al. 2013; Yang et al. 2017). To improve the quality of the genome assembly, the SGWG cultured an F_2 interspecific hybrid population derived from of the cross between *S. calycinum* ($2n = 32$, *var.* KEN8) and *S. angustifolium* ($2n = 32$, *var.* G01) and randomly chosen the 126 F_2 individuals from the population for genome re-sequencing on Illumina sequencing platform. Based on the Illumina data and the reference genomes of *S. calycinum* and *S. angustifolium*, the two interspecific maps for the wild *Sesamum* species were constructed for the first time. As a result, the SNP map of *S. calycinum* contained 1410 SNP markers located onto the 19 linkage groups (LGs), while the SNP map of *S. angustifolium* (*var.* G01) contained 1762 SNP markers onto the 22 LGs (Unpublished data, Haiyang Zhang).

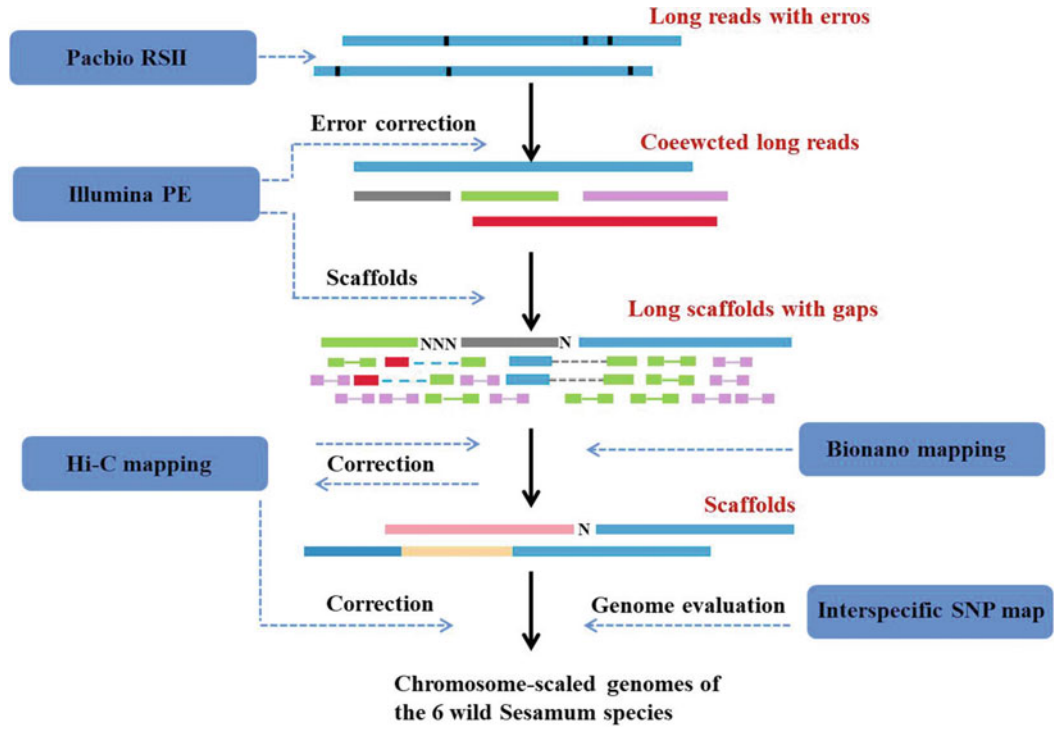


Fig. 17.1 Complicated genome sequencing and assembly strategies for the wild species. (Provided by Haiyang Zhang)

17.4 Genome Assembly of the Wild *Sesamum* Species

The chromosome-scaled genome sequences of the 6 wild *Sesamum* species were constructed for the first time. The assembled genomes ranged from 300.78 Mb (*S. calycinum*) to 625.03 Mb (*S. radiatum*) (Unpublished data, H. Zhang) (Table 17.2).

Considering the huge information, the genome analysis results of a representative species *S. radiatum* was introduced in this section, due to

Table 17.2 Size of the genome assemblies of the wild species and sesame

Sample no.	Species name	Genome size (Mb)
1	*S. indicum*	335.19
2	*S. alatum*	528.40
3	*S. latifollum*	369.09
4	S. angolense	312.97
5	*S. calycinum*	300.78
6	*S. angustifolium*	300.72
7	*S. radiatum*	625.03

Table 17.3 Distribution of interspersed repetitive sequences in in *S. radiatum* (var. G02)

TE type	TE number	Size (bp)	Percentage of genome (%)
SINEs	**7046**	**1,035,879**	**0.17**
ALUs	0	0	0.00
MIRs	0	0	0.00
Others	7046	1,035,879	0.17
LINEs	**12,314**	**12,012,151**	**1.92**
LINE1	10,641	11,733,431	1.88
LINE2	0	0	0.00
L3/CR1	0	0	0.00
Others	1673	278,720	0.04
LTR elements	**110,081**	**76,418,588**	**12.23**
DNA elements	**88,627**	**27,844,605**	**4.45**
hAT-Charlie	1470	181,420	0.03
TcMar-Tigger	0	0	0.00
Others	87,157	27,663,185	4.42
Unclassified	**660,628**	**158,461,945**	**25.35**
Total interspersed repeats	**878,696**	**275,773,168**	**44.12**

the high quality and the important phylogenetic position in *Sesamum*.

17.4.1 Distribution of the Interspersed Repetitive Sequences

To reveal the genome complexity and the ratio of the repetitive sequences in genome, the SGWG initially analyzed the repetitive sequence composition in the genome sequences of *S. radiatum* using RepeatMasker (version 4.0.6) (Maja and Nansheng 2009) and RepBase database (version 20,150,807). There are five groups of interspersed repetitive sequences including SINEs, LINEs, long terminal repeat (LTR) element, DNA element, and unknown dispersed repeat elements (Table 17.3). The results showed that the content of interspersed repeat sequences is 44.12%. Similar to the results of sesame, the LTR group is the biggest known repeat type with the number of 110,081, and occupies 76,418,588 bp (12.23%) of the assembled genome (625.03 Mb). Followed is the DNA elements group occupying 4.45% of the assembled genome. Compared with several genomes close to *Sesamum* in the phylogenetic tree of angiosperm plants, such as potato (*Solanum tuberosum*) (62.2% of the assembled 760 Mb genome) (The Potato Genome Sequencing Consortium, 2011), cacao (*Theobroma cacao*) (25.7% of the assembled 326.9 Mb genome) (Argout et al. 2011), grapevine (*Vitis vinifera*) (41.4% of the assembled 487 Mb genome) (Jaillon et al. 2007), the content of the repetitive/transposable elements (44.12%) of *S. radiatum* genome is moderate. The results infers that the genome of *S. radiatum* be uncomplicated.

17.4.2 Cetromeric Repeats in *S. radiatum*

Based on the *Si*Cen1 repeats of *S. indicum*, the SGWG screened the repeated DNA sequences of *S. radiatum* genome using RepeatMasker

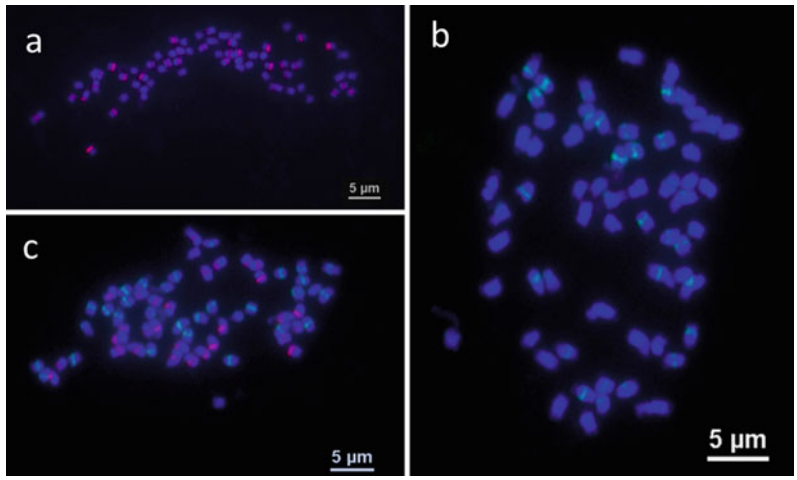


Fig. 17.2 FISH of *S. radiatum* using '*Sra*Cen1' and '*Sra*Cen2' repeat probes. **a** Fluorescence hybridization in situ using probe of '*Sra*Cen1' repeat (red signal) with about 16 chromosome pairs of *S. radiatum*. **b** Fluorescence hybridization in situ using probe of *Sra*Cen2 repeat (green signal) with about 16 chromosome pairs of *S. radiatum*. **c** Fluorescence hybridization in situ using two probes, '*Sra*Cen1' (red) and *Sra*Cen2 repeats (green). *Sra*Cen1 and *Sra*Cen2 repeats are labeled with tetramethyl-rhodamine and fluoresce in, respectively. Bar = 5 μm. (Provided by Haiyang Zhang)

Table 17.4 Distribution of *Sra*Cen1 and *Sra*Cen2 repeat sequences in *S. radiatum* genome

Chromosome no.	Total centromere repeat length (bp)	Chromosome no.	Total centromere repeat length (bp)
Chr.1	4003	Chr.17	11,891
Chr.2	258	Chr.18	694
Chr.3	0	Chr.19	2633
Chr.4	1716	Chr.20	910
Chr.5	11,089	Chr.21	11,987
Chr.6	14,183	Chr.22	271
Chr.7	1526	Chr.23	7255
Chr.8	0	Chr.24	733
Chr.9	0	Chr.25	2380
Chr.10	2833	Chr.26	4906
Chr.11	288	Chr.27	4349
Chr.12	0	Chr.28	18,396
Chr.13	14,335	Chr.29	10,341
Chr.14	3355	Chr.30	0
Chr.15	674	Chr.31	0
Chr.16	6079	Chr.32	6009
Unanchored scaffolds	1,666,487		
Total	1,809,581		

Note Unanchored scaffolds refer to the scaffolds unassembled onto the 32 chromosomes. (Provided by Haiyang Zhang)

(version 4.0.6). The threshold was at with following parameters: P < 1e-5, alignment length 120 bp, an identity ratio ≥ 90%. As a result, two repeat sequences were found in *S. radiatum*. We named these repeat sequences '*Sra*Cen1' (152 bp) and '*Sra*Cen2' (153 bp) as the candidate specific centromere repeats in *S. radiatum*. Interestingly, the fluorescence hybridization in situ (FIGH) (Zhao et al. 2018) of the two repeats as probes indicated that the two centromere repeats independently hybridized onto two equal groups of the 16 chromosome pairs of *S. radiatum* (Fig. 17.2). The results accorded with the cytological and genomehybridization results (shown in Chap. 9) that *S. radiatum* might be a polyploidy species.

Furthermore, the SGWG calculated the distribution of *Sra*Cen1 and *Sra*Cen2 repeats in the assembled genome of *S. radiatum* (Table 17.4). The total length of the *Sra*Cen1 and *Sra*Cen2 repeats was 1,809,581 bp, and most chromosomes carried the centromere repeats. The assembled centromere repeats in the genome supply the reliable information for the following chromosome group identification and genome evolution analysis of *S. radiatum*.

17.4.3 Distribution of Genes in *S. radiatum* Genome

Different from *S. indicum*, the wild species *S. radiatum* (var. G02) has 32 pairs of chromosome in somatic cell. The assembled genome of *S. radiatum* (var. G02) contains 73,126 genes, and 68,339 (93.45%) are localized onto the 32 chromosomes (Table 17.5).

Table 17.5 Distribution of genes in chromosomes of *S. radiatum*

Chromosome no.	Gene number	Gene length (bp)	Chromosome no.	Gene number	Gene length (bp)
*Sra*Chr.1	1662	3,858,358	*Sra*Chr.17	2127	4,345,713
*Sra*Chr.2	3279	7,527,304	*Sra*Chr.18	1764	3,841,659
*Sra*Chr.3	3417	7,449,515	*Sra*Chr.19	1942	3,892,704
*Sra*Chr.4	3168	7,105,475	*Sra*Chr.20	1935	4,731,676
*Sra*Chr.5	2809	6,918,489	*Sra*Chr.21	2265	5,631,480
*Sra*Chr.6	2602	5,215,618	*Sra*Chr.22	1590	3,485,100
*Sra*Chr.7	2220	4,559,885	*Sra*Chr.23	1712	3,561,510
*Sra*Chr.8	2327	4,981,269	*Sra*Chr.24	1917	3,596,086
*Sra*Chr.9	2382	5,240,068	*Sra*Chr.25	1635	3,759,476
*Sra*Chr.10	1748	4,045,310	*Sra*Chr.26	1930	3,825,200
*Sra*Chr.11	2165	4,982,624	*Sra*Chr.27	1906	4,269,671
*Sra*Chr.12	2500	5,794,374	*Sra*Chr.28	1979	3,983,214
*Sra*Chr.13	2133	4,410,779	*Sra*Chr.29	1738	3,758,563
*Sra*Chr.14	2262	4,591,349	*Sra*Chr.30	1586	3,239,686
*Sra*Chr.15	2109	4,380,012	*Sra*Chr.31	1588	3,534,002
*Sra*Chr.16	2043	3,990,539	*Sra*Chr.32	1899	4,156,844
Unanchored scaffolds				4787	5,203,223
Total				73,126	153,866,775

Note Unanchored scaffolds refer to the scaffolds unassembled onto the 32 chromosomes. (Supplied by Haiyang Zhang)

References

Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G et al (2011) The genome of *Theobroma cacao*. Nat Genet 43(2):101–108

Brennan AC, Barker D, Hiscock SJ, Abbott RJ (2014) Molecular genetic and quantitative trait divergence associated with recent homoploid hybrid speciation: a study of *Senecio squalidus* (Asteraceae). Heredity 108 (2):87–95

Doležel J, Binarová P, Lcretti S (1989) Analysis of nuclear DNA content in plant cells by flow cytometry. Biola Plant 31(2):113–120

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467

Joshi AB (1961) Sesamum. Indian Central Oilseed Committee Hyderabad, India, pp 1–3

Maja TG, Nansheng C (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics 25, Chapter 4: Unit 4.10

Miao HM, Li C, Duan YH, Wei LB, Ju M et al (2019) Identification of a Sidwf1 gene controlling short internode length trait in the sesame dwarf mutant dw607. Theor Appl Genet 133:73–86

Nimmakayala P, Perumal R, Mulpuri S, Reddy UK (2011) Sesamum. In: Kole C (ed) Wild crop relatives: Genomic and breeding resources. Volume: Oilseeds. Springer, Berlin, pp 261–273

Subramanian M (2003) Wide crosses and chromosome behavior in Sesamum. Madras Agric J 90:1–15

Yang M, Liu H, Zhou T, Qu H, Yang Y, et al. (2017) Production and identification of F1 interspecific hybrid between *Sesamum indicum* and wild relative *S. indicatum*. Sci Agri Sin 50(10):1763–1771

Zhang H, Miao H, Li C, Wei L, Ma Q (2012) Analysis of sesame karyotype and resemblance-near coefficient. Chin Bull Bot 47(6):602–614

Zhang H, Miao H, Zhang T, Wei L, Li C et al. (2013) Biological characters of interspecific hybrid progenies between *Sesamum indicum* L. and wild relatives (*Sesamum schinzianum* Asch, *Sesamum radiatum* Schum &Thonn). Sci Agri Sin 46 (19):3965–3977

Zhang HY, Miao HM, Ju M (2019) Potential for adaptation to climate change through genomic breeding in sesame. In: Kole C (ed) Genomic designing of climate-smart oilseed crops. Springer, Cham, Switzerland, pp 374–376

Zhao RH, Miao HM, Ma Q, Chen CB, Song WQ et al (2018) Karyotype comparison analysis of the wild species *Sesamum alatum* and the cultivated sesame. Acta Sci Natural Univ Nankaiensis 5(05):27–36

The Sesame Genome for Gene Discovery in Sesame

18

Yinghui Duan, Ming Ju, Hongmei Miao, and Haiyang Zhang

Abstract

Implementation of the Sesame Genome Project and the achievement of sesame genome research provide abundant bioinformatics information for the theoretical and applied genetic analyses of the important agronomic traits and biological processes in sesame. In order to facilitate the application of the genome information and bioinformatics, we present the discovery of some key genes such as *SiDt1*, *Sidwf1*, *SiOPP*, and *Sicl1* and describe the main methods of gene cloning based on the sesame genome data. The potential of gene digging using the high-quality assembled genome is also discussed in this chapter.

18.1 Introduction

Sesame is cultivated widely in the tropical and subtropical regions of Asia, Africa, and South America (Zhang et al. 2019). Considering the species characters and the cultivation environment conditions, the main objectives of sesame breeding in the world are always involved in the yield potential and the resistance and tolerance to biotic and abiotic stresses (Ashri 2006). In the past decade, construction of several reference genomes for *S. indicum* (Zhang et al. 2013; Wang et al. 2014, 2016; Kitts et al. 2016) and the wild *Sesamum* species (the Sesame Genome Working Group, 2010–2020, H. Zhang) improved the sesame genomics research and finally stimulated sesame breeding ultimately. For instance, based on the chromosome-scaled reference genome, a *SiDt* gene regulating the inflorescence meristem determinacy in sesame has been cloned for the first time, which explains the phenomena of eternal determinacy and low capsule node number (less than five) in *dt2* type germplasm (Zhang et al. 2016). Till now the *dt1* gene marker has been applied in selection of determinate varieties and evaluation of seed purity (Data not shown). Moreover, with the aid of the Illmunia de novo genome assembly information, some candidate genes such as *SiACS* controlling the flower number per axil and *SiPPO* gene regulating the seed coat color were determined using the genome-wide association analysis (GWAS) of the 705 sesame accessions (Wei et al. 2015). At present, dozens of genes and gene markers have been determined and some have been applied for variety breeding (Zhang et al. 2019). Thus, we here summarize the main methods and strategies of gene cloning

Y. Duan · M. Ju · H. Miao · H. Zhang (✉)
Henan Sesame Research Center, Henan Academy of Agricultural Sciences, Zhengzhou, People's Republic of China
e-mail: zhanghaiyang@zzu.edu.cn

H. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-98098-0_18

using the reference genome to facilitate the application of the genome information in large-scaled gene digging in future.

18.2 Gene Discovery via Traditional Homolog Gene Cloning Methods

Before initiation of the Sesame Genome Project, traditional homologous gene cloning was the main method for sesame gene analysis. In 2003, the first sesame transcriptome data, 3328 expressed sequence tags (ESTs) were obtained from the developing seeds (Suh et al. 2003). Subsequently, more transcriptome and ESTs data were obtained and applied to determine the transcriptional structure of genes (Wei et al. 2009, 2011). As a result, a few functional genes regulating the formation and regulation of lipids, storage proteins, and secondary metabolites and the response to salt stress in sesame are firstly cloned with the aid of homologs'information and transcriptome data (Yukawa et al. 1995; Chen et al. 1997; Jin et al. 2001; Chyan et al. 2005; Lee et al. 2006; Hsiao et al. 2006; Kim et al. 2007, 2010; Hata et al. 2010).

18.2.1 Detection of SeCKI-SebHLH-SeFAD2 Pathway Using Homolog Amplification

As sesame is an oilseed crop with high oil content and quality, sesame scientists pay more attention to the key genes regulating the lipid synthesis and metabolism (Tzen 2012; Tzen et al. 1993, 2003). Desaturation catalyzed by ω-6 fatty acid desaturases is a biochemical process to desaturate oleic acid to linoleic acid. In plants, the genes for ER- and plastid-derived ω-6 fatty acid desaturases are *fad*2 and *fad*6, respectively. Based on the homolog sequence information and a sesame cDNA library, the cDNA sequence of ω-6 fatty acid desaturase in sesame was obtained from developing seeds using RT-PCR (Jin et al. 2001). The cDNA sequence was named *SeFAD*2 of 1466 bp. Further function analysis indicated that *SeFAD2* transcripts could be regulated by abscisic acid (ABA) in developing sesame seeds. Moreover, two genes *SebHLH* and *SeCKI* regulating *SeFAD*2 function were cloned based on transcriptome information (Kim et al. 2007). The results revealed that SeCKI enhanced the SebHLH-mediated transactivation of the *SeFAD2* gene promoter and determined the function of the SeCKI-SebHLH-SeFAD2 pathway in lipid biosynthesis in sesame seeds.

18.2.2 CYP81Q1 Gene Discovery via Homolog Function Analysis

Sesame lignans are a group of antioxidants in sesame seeds and leaves, and are responsible for health benefit for human beings (Shittu et al. 2007). Sesamin and sesamolin are the main components of lingans. Ono et al. (2006) found that sesamin is synthesized from (+)-pinoresinol which is finally mediated by two distinct cytochrome P450 proteins (i.e., SiP450 and CYP81Q1) in sesame. The CYP81Q1 gene was expressed differentially in different sesame genotypes and affected the final sesamin content in leaves (Hata et al. 2010; Naoki et al. 2012). In sesame accession CO-1, the transcript abundance of *CYP81Q1* gene generated along with high sesamin content, similar to the wild specie *S. malabaricum* (Pathak et al. 2015).

18.2.3 *SeMIPS1* Gene Isolation from Transcriptome via Homolog Analysis

Tolerance to abiotic stresses is an important breeding objective in sesame. Considering myo-inositol 1-phosphate synthase (MIPS) might play a role in the response to salinity stress, Chun

et al. (2003) isolated the *SeMIPS1* gene from sesame seeds. MIPS protein catalyzes glucose-6-phosphate to myo-inositol 1-phosphate, which is the first product in the biosynthetic pathways of myoinositol, phytic acid, and other essential cellular components (Loewus et al. 1990; Yoshida et al. 1999; Loewus and Murthy 2000). Comparison results showed that SeMIPS1 protein was highly homologous with those from other plant species (88–94%). In the sesame genome, several copies of the *SeMIPS1* gene have been detected (data not shown). Under the increased saline conditions, transcription of SeMIPS1 in germinating seeds was greatly reduced, indicating the related response to salt stress in sesame (Chun et al. 2003).

18.3 Gene Cloning via Fine Linkage Mapping

With the rapid development of next-generation sequencing (NGS) technologies, new strategies of gene cloning are proposed and applied in sesame (Zhang et al. 2019). Based on the high-quality reference genome, the first high-dense single nucleotide polymorphism (SNP) genetic linkage map was constructed and used for *SiDt* gene location in sesame in 2016 (Zhang et al. 2016). Linkage mapping analysis is widely used for gene discovery and molecular marker location.

By the early stage of the Sesame Genome Project (SGP), the Sesame Genome Working Group (SGWG) constructed an ultra-dense SNP genetic linkage map using whole-genome re-sequencing platform for guiding the reference genome assembly (Zhang et al. 2013, 2016). Later, the SNP genetic map was applied to clone the *SiDt* gene because of the various determinacy traits in the two parents. For Of the two parents of the assayed F_2 population, JS012 (*Dt*, P_2) is an indeterminate variety with normal inflorescence meristem (Fig. 18.1), while cv. Yuzhi DS899 (*dt1*, P_1) is a determinate mutant induced from ethylmethane sulfonate (EMS) mutagenesis. The determinate parent could develop 8–20 capsule nodes, and the number of capsule nodes varies with the latitude change. Interestingly, the *dt*1-type presented in Yuzhi DS899 (*dt1*) is a different allele from the *dt2* type (08TP092) which was induced by gammar induction (Ashri 2001) and

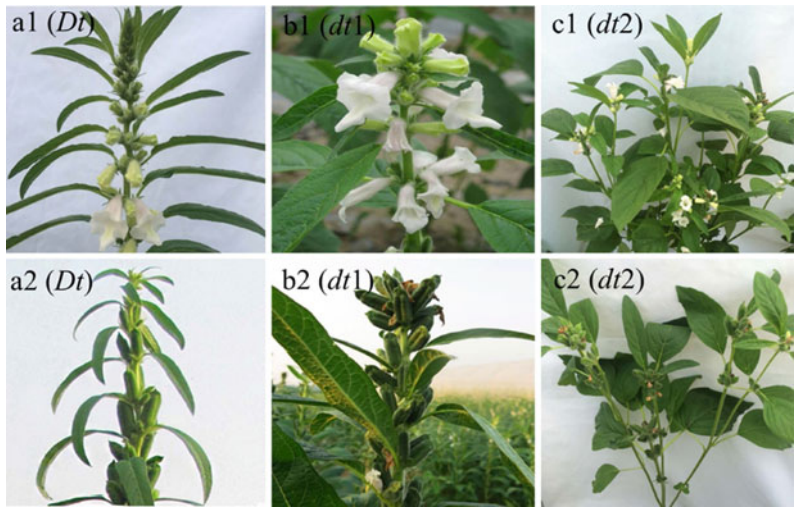


Fig. 18.1 Phenotypes of inflorescence meristem development in sesame. **a1** and **a2**: the indeterminate genotype (P1, *Dt*) (the wild type). **b1** and **b2**: Yuzhi DS899, determinate type 1 genotype (*dt1*) with a terminal flower (or capsule) cluster at shoot apex. **c1** and **c2**: 08TP092, determinate type 2 genotype (*dt2*) with a terminal flower (or capsule) cluster at apex of each shoot. Only 2–3 capsule nodes form on each branch. (Cited from Zhang et al. (2016))

the wild type (P_1). Genetic background analysis of all F_1, BC_1 and F_2 generations from direct crosses and reciprocal crosses between *Dt* and *dt* types validated that *dt2* (08TP092) was allelic with *dt1* (Yuzhi DS899) and controlled by the single gene locus.

With aid of the genome re-sequencing data of the 120 F_2 progeny and the parents, the SGWG constructed an ultra-dense SNP genetic linkage map using JoinMap software. A total of 3041 bin markers containing 30,193 SNP markers were mapped on 13 linkage groups. The ultra-dense SNP map was 2981.28 cM in length. The average marker density was saturated with approximately 0.98 cM per bin or 0.10 cM per SNP and could be applied for genome assembly and gene location. Based on the ultra-dense SNP map and phenotypic segregation of the F_2 population, gene locus mapping was carried out using winQTLcart and QTLNetwork, respectively. The QDt1 locus for the determinacy trait in sesame was determined in the 18.0–19.2 cM. In the reference genome, the QDt1 covered 292.35 kb sequences which comprised of 27 predicted genes in Scaffold 00,170. Further screening of SNPs and InDels in the 292.35 kb fragment showed that only the SNP marker *Si*Dt27-1 entirely co-segregated with the determinate trait. The marker existed in the target gene DS899s00170.023 named *SiDt1* (Fig. 18.2) (shown in the Chap. 8).

Gene sequence comparison revealed that Si*Dt* gene is comprised of 4 exons and 3 introns with 531 amino acid residues (Fig. 18.2). A SNP of G397A resulted in the difference of *Sidt1* from *SiDt* genes and caused the change of the 79th amino acid residue from S to N. *SiDt* gene in sesame was high homology with the TFL gene of *A. thaliana, which* controls the morphological switch between shoot growth and the inflorescence development (Bradley et al. 1997). According to 08TP092 genome data (PRJNA316751), *dt*2 type lost the whole *SiDt* gene sequence and presented the eternal determinacy. The results reflect the important function of *SiDt* in regulating the inflorescence meristem development pathway in sesame. At present, the SNP marker has been applied for screening of new varieties with the determinate habit trait (Zhang et al. Chinese patent no. ZL201510876016.3; USA Patent no. 10301687). All the above research suggests that map based

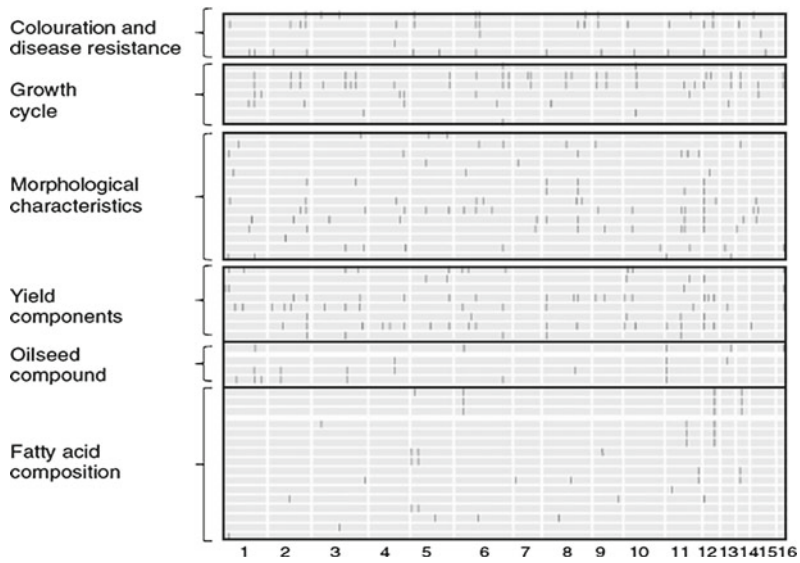


Fig. 18.2 SNP detection for 56 agronomic traits in sesame using GWAS analysis. (Modified from Wei et al. (2015))

gene cloning with the high-quality reference genome and linkage map is an effective to dig genes with major effect on target traits.

18.4 Gene Digging via Genome Wide Association Studies

Besides map-based gene cloning, GWAS analysis is also an effective method to determine the gene/quantitative trait locus (QTL) with major effect on the target traits. Based on linkage disequilibrium (LD) between molecular markers and complex traits, association analysis could be used to analyze the relationship between genotype variation and phenotype variation.

The development of sequencing technologies and coupled computation methods significantly accelerates construction of high-density haplotype maps, which comprehensively capture the genomic variation and the pattern of common haplotypes within the species (Atewll et al. 2010; Huang et al. 2010, 2011). To our knowledge, the genomes of more than 1500 cultivated sesame accessions have been re-sequenced using the Illumina HiSeq2000 or HiSeq2500 system till now. By exploiting the sequence variation in the genomes of the nature populations, the (GWAS) of 705 sesame germplasm accessions in sesame could be performed for multiple agronomic traits for the first time (Wei et al. 2015). The genomes of 705 accessions were re-sequenced using the Illumina HiSeq2500 system, each with ~ 2.6-fold genome coverage. The sesame cultivar 'Zhongzhi no. 13' genome sequence was used as the reference (Wang et al. 2014). A total of 5,407,981 SNPs were identified in the population. As a result, a total 549 loci were identified associated the phenotypic variation of the 56 agronomic traits (Fig. 18.2).

Based on the major variant loci, 46 candidate genes were believed related to oil content, fatty acid biosynthesis, yield, and other traits in sesame. As to the GWAS analysis, the *SiACS* gene (SIN_1006338) controlling the flower/capsule number per axil was determined with the explanation ratio up to 60% to the phenotypic variation. In the *SiACS* gene, a missense SNP within *SiACS* could lead to the amino acid variation of F284S. *SiACS* is homologous with AtACS8 in Arabidopsis which was reported involved in ethylene biosynthesis (Tsuchisaka and Theologis 2004). Notably, as to the complex traits, the candidate genes dig from the GWAS analysis need more validation. To ensure the effects of the candidate genes and the variants on the associations underlying the important traits, more functional genomics studies such as genetic transformation and genome-editing technology using CRISPR/Cas system should be performed. On the other hand, construction of multiple biparental populations from the well-designed crosses would determine the target genes and identify the epistatic interactions of multiple genes.

18.5 Gene Detection via Cross-Population Association Mapping and Genomic Variants Screening

In recent years, integration of linkage mapping and association mapping have been successfully applied for detection of QTLs and candidate genes linked to specific complex agricultural traits in many crops (Cadic et al. 2013; Korir et al. 2013; Zhao et al. 2014; Shi et al. 2015; Deborah et al. 2017). In order to identify the target gene regulating the curly leaf and close capsule traits in sesame, the cross-population association mapping and genomic variants screening method was applied in sesame for the first time (Zhang et al. 2018). A total of 130 F_2 individuals derived from a cross between the curly leaf mutant *cl1* and the wild type parent was sequenced using Illumina sequencing platform. Genetics analysis of the mutagenesis character in the population proved that the curly leaf trait of sesame is controlled by single gene pair.

For the 130 individuals of the F_2 population, a total of 425,661 variants were plotted in the 13 chromosomes. Accurate association analysis of the variants indicated that a ~ 400 kb region with significantly high *P* value was located in

LG8. Differing from the traditional linkage mapping method, cross-population association mapping analysis was performed using GLM (general linear model) model in TASSELE5.0. The fine sesame genome (var. Yuzhi 11, GCA_001692995.1) and the genome re-sequencing variants data supplied the reliable genomic information for gene detection (Zhang et al. 2018). As a result, the *SiCL1* gene controls leaf curling and capsule indehiscence in sesame was cloned (Fig. 18.3). *SiCL1* gene encodes a transcription repressor KAN1 protein. In *cl1* mutant, the 20 nucleic acids (CAGGTAGCTATGTATATGCA) of *SiCLInDel1* marker were mutagenized into 6 nucleic acids (TCTTTG), and caused a frameshift mutation and finally resulted in the earlier translation termination of the CL gene. The findings provided an example of high-efficient gene cloning method in sesame (Zhang et al. 2018).

18.6 Dene Discovery via MutMap Analysis

Bulked segregant analysis (BSA) provides a simple and effective alternative strategy for identifying molecular markers linked to target genes by genotyping only a pair of bulked DNA samples from two sets of individuals with distinct or extremely opposite phenotypes (Michelmore et al. 1991). With the applications of whole-genome sequencing (WGS), BSA strategies have been optimized and applied for identifying QTLs or candidate genes (Abe et al. 2012; Takagi et al. 2013). Abe et al. (2012) introduced MutMap, a method based on whole-genome re-sequencing of pooled DNA from a segregating population of plants, and identified the unique variant causing pale green leaves and semi dwarfism trait. Takagi et al. (2013) identified the QTLs associated with the key agronomic traits such as partial resistance to the fungal rice blast disease and seedling vigor in rice by whole-genome re-sequencing. Recently, the SGWG-HSRC (the SGWG group member, Henan Sesame Research Center, Henan Academy of Agricultural Sciences, China) applied the Mutmap method in the BSA analysis using 54 individuals of a F_2 population derived from a cross between a susceptible mutant (S) to SFW and the wild type (R) with high resistance to SFW (Data not shown) (Unpublished data, Haiyang Zhang). Finally, a RLK gene controlling the resistance to *Fusarium* wilt disease was cloned and proved the high efficiency of MutMap method using the BSA strategy.

18.7 Dene Discovery via Comparative Genomics Analysis

Similar to detecting the homolog genes, comparative genomics analysis realizes the detection of genes or gene family of unknown species, based on the similarity of the arrangement or structure of chromosomal or orthologous of gene in many known species genome. As shown in the former chapters about gene families annotation,

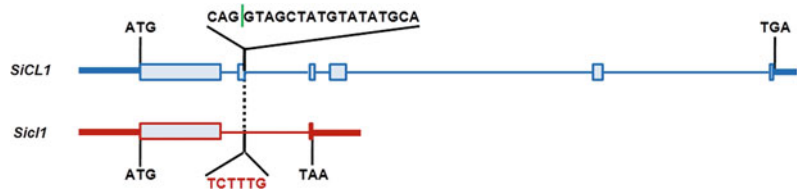


Fig. 18.3 Cloning of SiCL1 controlling the curly leaf and indehiscent capsule. In the mutant *cl1*, The 20 nucleic acids (CAGGTAGCT ATGTATATGCA) in *SiCL1* are mutated into 6 nucleic acids (TCTTTG) which causes a frameshift mutation and an early translation termination. (Cited from Zhang et al. (2018))

some transcription factors families such as the MADS-box gene family, AP2/ERF transcription factor gene family, and Hsfs (Heat shock transcription factors) gene family in sesame have been analyzed, combined with genome comparison of sesame with including *Arabidopsis thaliana*, *Vitis vinifera, Solanum lycopersicum, Solanum tuberosum*, and *Utricularia gibba* (Wei et al. 2015; Dossa et al. 2016a, 2016b). However, more functional genomics studies should be performed to elucidate the functions of the orthologous gene families in sesame. Notably, all the above methods and strategies for gene discovery should utilize the reference genome and the genomes of the population as the basic information. Thus the high quality of a reference and abundant genome variants data and precise phenotypes of the population are emphasized on the success of gene discovery.

References

Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H et al (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. Nat Biotechnol 30:174–178

Ashri A (2001) Induced mutations in sesame breeding. No. IAEA-TECDOC-1195

Ashri A (2006) Sesame (*Sesamum indicum* L.). In: Signh RJ (ed) Genetic resources, chromosome engineering, and crop improvement. CRC Press, Boca Raton, FL, pp 231–280

Atewll S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M et al (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature 465(7298):627–631

Bradley D, Ratcliffe O, Vincent C, Carpenter R, Coen E (1997) Inflorescence commitment and architecture in Arabidopsis. Science 275:80–83

Cadic E, Coque M, Vear F, Grezes-Besset B, Pauquet J et al (2013) Combined linkage and association mapping of flowering time in sunflower (*Helianthus annuus* L.). Theor Appl Genet 126(5):1337–1356

Chen JCF, Lin RH, Huang HC, Tzen JTC (1997) Cloning, expression and isoform classification of a minor oleosin in sesame oil bodies. J Biochem 122:819–824

Chun JA, Jin UH, Lee JW, Yi YB, Hyung NI et al (2003) Isolation and characterization of a myo-inositol 1-phosphate synthase cDNA from developing sesame (*Sesamum indicum* L.) seeds: functional and differential expression, and salt-induced transcription during germination. Planta 216:874–880

Chyan CL, Lee TTT, Liu CP, Yang YC, Tzen JTC et al (2005) Cloning and expression of a seed-specific metallothionein-like protein from sesame. Biosci Biotechnol Biochemy 69(12):2319–2325

Deborah DA, Vemireddy LR, Roja V, Patil S, Choudhary GP et al (2017) Molecular dissection of QTL governing grain size traits employing association and linkage mapping in basmati rice. Mol Breed 37(6):77

Dossa K, Wei X, Li D, Fonceka D, Zhang Y et al (2016b) Insight into the ap2/erf transcription factor superfamily in sesame and expression profiling of dreb subfamily under drought stress. BMC Plant Biol 16 (1):171

Dossa K, Diouf D, Cissé N (2016a) Genome-wide investigation of Hsf genes in sesame reveals their segmental duplication expansion and their active role in drought stress response. Front Plant Sci 7:1522

Hata N, Hayashi Y, Okazawa A, Ono E, Satake H, et al. (2010) Comparison of sesamin contents and cyp81q1 gene expressions in aboveground vegetative organs between two Japanese sesame (*Sesamum indicum* L.) varieties differing in seed sesamin contents. Plant Sci 178(6):510–516

Hsiao ESL, Lin LJ, Li FY, Wang MMC, Liao MY et al (2006) Gene families encoding isoforms of two major sesame seed storage proteins, 11S globulin and 2S albumin. J Agri Food Chem 54:9544–9550

Huang X, Wei X, Sang T, Zhao q, Feng Q et al (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42(11):961

Huang X, Zhao Y, Wei X, Li C, Wang A et al (2011) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat Genet 44(1):32

Jin UH, Lee JW, Chung YS, Lee JH, Yi YB et al (2001) Characterization and temporal expression of a ω-6 fatty acid desaturase cDNA from sesame (*Sesamum indicum* L.) seeds Plant Sci 161:935–941

Kim MJ, Kim JK, Shin JS, Suh MC (2007) The sebhlh transcription factor mediates trans-activation of the sefad2 gene promoter through binding to e- and g-box elements. Plant Mol Biol 64:453–466

Kim MJ, Go YS, Lee SB, Kim YS, Shin JS et al (2010) Seed-expressed casein kinase i acts as a positive regulator of the sefad2 promoter via phosphorylation of the sebhlh transcription factor. Plant Mol Biol 73 (s4–5):425–437

Kitts PA, Church DM, Françoise TN, Jinna C, Vichet H et al (2016) Assembly: a resource for assembled genomes at NCBI. Nucleic Acids Res 2016(D1):D73–D80

Korir PC, Zhang J, Wu K, Zhao T, Gai J (2013) Association mapping combined with linkage analysis for aluminum tolerance among soybean cultivars released in Yellow and Changjiang River Valleys in China. Theor Appl Genet 126(6):1659

Lee TTT, Leu WM, Yang HH, Chen BCM, Tzen JTC (2006) Sesame oleosin and prepro-2s albumin expressed as a fusion polypeptide in transgenic rice

were split, processed and separately assembled into oil bodies and protein bodies. J Cereal Sci 44(3):333–341
Loewus FA, Murthy PPN (2000) myo-Inositol metabolism in plants. Plant Sci 150:1–19
Loewus FA, Everand JD, Young KA (1990) Inositol metabolism: precursor role and breakdown. In: Morre´ DJ, Boss WF, Loewus FA (eds) Inositol metabolism in plants. Wiley-Liss, New York, pp 21–45
Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88:9828–9832
Naoki H, Yoshinori H, Atsushi O, Eiichiro O, Honoo S et al (2012) Effect of photoperiod on growth of the plants, and sesamin content and cyp81q1 gene expression in the leaves of sesame (*Sesamum indicum* L.). Environ Exp Bot 75:212–219
Ono E, Nakai M, Fukui Y, Tomimori N, Fukuchi-Mizutani M et al (2006) Formation of two methylenedioxy bridges by a Sesamum cyp81q protein yielding a furofuran lignan, (+)-sesamin. Proc Natl Acad Sci USA 103(26):10116–10121
Pathak N, Bhaduri A, Bhat KV, Rai AK, Bekker R (2015) Tracking sesamin synthase gene expression through seed maturity in wild and cultivated sesame species—a domestication footprint. Plant Biol 17(5):1039–1046
Shi J, Zhan J, Yang Y, Ye J, Huang S et al (2015) Linkage and regional association analysis reveal two new tightly-linked major-QTLs for pod number and seed number per pod in rapeseed (*Brassica napus* L.). Sci Rep 5:14481
Shittu LAJ, Bankole M, Ahmed T, Bankole MN, Shittu RK et al (2007) Antibacterial and antifungal activities of essential oils of crude extracts of *Sesamum radiatum* against some common pathogenic microorganisms. Iran J Pharmacol Ther 6(6):165–170
Suh MC, Kim MJ, Hur CG, Bae JM, Park YI et al (2003) Comparative analysis of expressed sequence tags from *Sesamum indicum* and *Arabidopsis thaliana* developing seeds. Plant Mol Biol 52:1107–1123
Takagi H, Abe A, Yoshida K, Kosugi S, Natsume et al (2013) QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J 74:174–183
Tsuchisaka A, Theologis A (2004) Heterodimeric interactions among the 1-amino-cyclopropane-1-carboxylate synthase polypeptides encoded by the Arabidopsis gene family. Proc Natl Acad Sci USA 101(8):2275–2280
Tzen JTC (2012) Integral proteins in plant oil bodies. ISRN Botany Article ID 263270
Tzen JTC, Cao YZ, Laurent P, Ratnayake C, Huang AHC (1993) Lipids, proteins, and structure of seed oil bodies from diverse species. Plant Physiol 101:267–276
Tzen JTC, Wang MMC, Chen JCF, Lin LJ, Chen MCM (2003) Seed oil body proteins: oleosin, caleosin, and steroleosin. Curr Topics Biochem Res 5:133–139
Wang L, Yu S, Tong C, Zhao Y, Liu Y et al (2014) Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. Genome Biol 15 (2):R39
Wang J, Yu J, Sun P, Li Y, Xia R et al (2016) Comparative genomics analysis of rice and pineapple contributes to understand the chromosome number reduction and genome changes in grasses. Front Genet 7:174
Wei LB, Zhang HY, Zheng YZ, Miao HM, Zhang TZ et al (2009) A genetic linkage map construction for sesame (*Sesamum indicum* L.). Genes Genom 31 (2):199–208
Wei WL, Qi X, Wang LH, Zhang YX, Wei H et al. (2011) Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using illumina paired-end sequencing and development of est-ssr markers. BMC Genomics 12(1):451
Wei W, Zhang Y, Wang L, Li D, Gao Y et al (2015) Genetic diversity, population structure, and association mapping of 10 agronomic traits in sesame. Crop Sci 56(1). https://doi.org/10.2135/cropsci2015.03.0153
Yoshida KT, Wada T, Koyama H, Mizobuchi-Fukuoka R, Naito S (1999) Temporal and spatial patterns of accumulation of the transcript of myo-inositol 1-phosphate synthase and phytin containing particles during seed development in rice. Plant Physiol 119:65–72
Yukawa Y, Takaiwa F, Yoshida N, Yamada K (1995) Structure and expression of two seed –specific cDNA clones encoding steroyl-acyl carrier protein desaturase from sesame, *Sesamum indicum* L. Plant Cell Physiol 37:201–205
Zhang H, Miao H, Wang L, Qu L, Liu H et al (2013) Genome sequencing of the important oilseed crop *Sesamum indicum* L. Genome Biol 14(1):401–409
Zhang H, Miao H, Li C, Wei L, Duan Y et al (2016) Ultra-dense SNP genetic map construction and identification of SiDt gene controlling the determinate growth habit in *Sesamum indicum* L. Sci Rep 6:31556
Zhang H, Miao H, Wei L, Li C, Duan Y et al (2018) Identification of a SiCL1 gene controlling leaf curling and capsule indehiscence in sesame via cross-population association mapping and genomic variants screening. BMC Plant Biol 18(1):296
Zhang H, Miao H, Ju M (2019) Potential for adaptation to climate change through genomic breeding in sesame. In: Kole C (ed) Genomic designing of climate-smart oilseed crops. Springer, Cham, Switzerland, pp 374–376
Zhao Y, Wang H, Chen W, Li Y (2014) Genetic structure, linkage disequilibrium and association mapping of *Verticillium* wilt resistance in elite cotton (*Gossypium hirsutum* L.) germplasm population. PLoS One 9(1): e86308

Prospect of Designed Breeding in Sesame in the Post-genomics Era

19

Haiyang Zhang, D. Ray Langham, and Hongmei Miao

Abstract

The achievement of the Sesame Genome Project supplies the necessary information for sesame genome assisted selection breeding. The information of the structural genomics and functional genomics research in sesame provides the foundation of genome architecture analysis. Cloning of many genes and gene families makes the possibility of aggregating a number of elite traits in new varieties. In order to enhance the variety improvement, the specific genome assisted selection breeding technology for sesame is designed and discussed in this chapter. The potential of the genome assisted selection and gene design breeding in sesame is forecasted accordingly.

19.1 Introduction

In 1984 at the FAO Viterbo, Italy meetings of sesame and safflower experts, the experts predicted that if sesame were not mechanized in the next 25–30 years, its world production would decrease significantly. Without mechanization, sesame would only persist in those niches where no other suitable crop could be grown. Thirty-three years after the prediction, the production of sesame worldwide is still increasing. Although sesame is completely mechanized in the United States, the production is less than 1% of the world production. There has been some increasing mechanization in Paraguay using the Venezuela technology. However, almost all the sesame produced still is dependent on at least some manual labor. The *mantra* has always been that labor is cheap for sesame in the developing countries. However, the challenges of low profit and high planting cost still exist and affect the development of the world sesame industry. On the other hand, the hindrance of sesame variety improvement also requires the sesame breeders to develop new breeding techniques and breeding materials to meet with the breeding objectives and the industry development rhythm. As the Sesame Genome Project (SGP) initiated in 2010, the Sesame Genome Working Group (SGWG) planned to elucidate the phylogenetic analysis of the cultivated species in *Sesamum* and the regulation of the key traits such as yield, seed quality,

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

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

H. Miao et al. (eds.), *The Sesame Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-98098-0_19

and resistance to the main diseases with the aid of the genome sequence of *S. indicum* (var. Yuzhi 11) (Zhang et al. 2013a). Comparative genomics of the sesame genome with the genomes of monocotyledonous and other dicotyledonous plants indicated that both the genome architecture and the gene families take part in the regulation of the important biological processes in sesame. As shown in the former chapters, the application of the precise genome sequences is accelerating sesame genomics and genetic research and should improve sesame breeding in the future.

19.2 Ideal Sesame Varieties and Breeding in the Future

As to the world sesame production, farmers need to have the highest profit as long as their crop rotations are sustainable. Thus, the ideal sesame varieties should have high yield potential and stable adaptability to the environmental conditions. In the past decades, the main objectives of sesame breeding focused on breeding new varieties with high yield level and the high resistance/tolerance to diseases, pests, waterlogging, drought, and low temperature stresses (Ashri 2006). However, the future of sesame is unpredictable and affected by the industry development. Although sesame yields have been improved, the increases are not as high as the other crops. Thus, sesame farmers are switching to the more profitable crops. Meanwhile, breeders are hesitant to convert sesame to a genetically-modified (GM) crop because the market wants sesame to stay non-GMO (non-genetically-modified organism). If sesame cultivation is not mechanized, it will be limited to areas that are not mechanized and have no other crop options. The high demand for sesame in China, Japan, South Korea, and India will provide higher prices, but higher prices may lead to less consumption if the price is out of the range of the common household. Regardless of mechanization, unless breeders and agronomists improve yields, sesame will not compete with other crops. However, sesame seeds contain more than 50% oil and 25% protein with antioxidants that can be eaten as is, toasted, crushed for oil, made into *tahini*, and used medicinally. The lore that has been passed down for generations is 'Eat sesame is for health', and science continues to show that in fact sesame does improve heath. The future of sesame is in the breeders' and agronomists' hands.

As to sesame breeding technology, genomics is the key to reducing the breeding time from conception to reality. Although sesame is lagging behind the major crops in genomics and the new gene technologies to improve crops, the 'genome time' for sesame is coming. Whether it is 10, 20, 30, or 40 years away, the time will come that we need to think about the ideal sesame. It will not be necessary for each research institute to acquire all the equipment and expertise. Entrepreneurs will build companies that will move a trait or multiple traits for a cost far below developing in-house genomics. Scientists will have the ability to assemble genomes, and breeders need to define the genome to be assembled. Someone in Turkey will find a cultivar that is tolerant to phyllody; the responsible gene(s) will be identified; the gene(s) will be added to a cultivar in Pakistan. Classical breeding is a long, long, long process. The normal cycle from cross to variety is 7–10 years. Most of the time multiple crosses must be made to move one trait to an acceptable genotype. The progenitor of all non-dehiscent sesame was discovered in an F_2 against odds over 1 in 1,000,000 plants. With the genomic selection, sesame breeders could easily choose the target plant with the ideal traits from the large population in the future.

19.3 Potential of Genomics-Assisted Breeding in Sesame

Genome-assisted breeding refers to the application of molecular markers associated with important traits and the genomic tools to predict the phenotype and to assist breeding (Varshney et al. 2013). Before the initiation of the SGP, only a few genes related to lipid formation and regulation, seed storage proteins, and regulation

of secondary metabolites, and salt stress response have been cloned (Yukawa et al. 1995; Chen et al. 1996; Jin et al. 2001; Chyan et al. 2005; Lee et al. 2006; Hsiao et al. 2006; Kim et al. 2007; Hata et al. 2010; Kim et al. 2010). With the aid of high through-put next-generation sequencing (NGS) and the third sequencing technologies and the precise genome information, the genetics and genomics research in sesame was extremely improved in recent a few years (Zhang et al. 2013a,b; Miao 2014; Wang et al. 2014,2016; Wei et al. 2015; Miao and Zhang 2016; Yu et al. 2019). Hundreds of candidate genes, quantitative trait loci (QTLs), and molecular markers associated with dozens of agronomic traits help the breeders to know the genetics basis of the target traits and materials (Wei et al. 2015; Dossa et al. 2016; Zhang et al. 2019).

In addition, genome re-sequencing data of the natural population reflect the distribution of the DNA fragment linkages in sesame genome (Unpublished data, Haiyang Zhang) and give the guidance for estimating the selection efficiency of hybridization of a specific cross with target traits sesame (Zhang et al. 2013b; Wei et al. 2015; Wang et al. 2016; Zhang et al. 2016). Based on the ultra-density single nucleotide polymorphism (SNP) genetic map (Zhang et al. 2016) and the updated genome sequence, 13 elite gene clusters related to the 58 markers ($P < 10^{-8}$) and 52 QTLs ($R^2 > 10\%$) associated with yield, seed quality and *Fusarium* wilt disease resistance traits were detected in eight chromosome pairs of *S. indicum* (Miao and Zhang 2016) (partially unpublished, Haiyang Zhang). In 2016, based on the reference genome, the SGWG-HSRC cloned the *SiDt* gene controlling the determinacy of the inflorescence meristem development and the SNP marker was applied in evaluating the purity of the determinate variety seeds. The findings of the genetic mapping and association mapping related to the key agronomic traits will reinforce the molecular breeding in sesame. We believe that the molecular breeding system would be constructed in near future.

19.4 Genomics-Assisted Breeding Routine for Sesame

In recent few years, molecular genetics and genomics research in sesame was accelerated. However, there are no reports about genomics-assisted breeding in sesame. DNA-based molecular markers have the genetic stability and the advantages for selection in crop breeding (Gupta et al. 2010; Kumpatla et al. 2012). Introducing and applying molecular breeding methods to improve the efficiency of breeding programs is necessary.

In the genomics-assisted breeding system, precisely selecting the desirable and elite traits is very important for the plant breeders, even though most of the agronomic traits are controlled by polygenes with complex nonallelic quantitative effects. In order to establish the relationship between genotype and phenotype, the genomics and genetics analyses of the target traits are necessary. Differing from the conventional breeding procedures, the molecular maker-assisted breeding system with five major steps is designed for sesame molecular breeding (Fig. 19.1).

The five major steps are: (1) Create and screen more elite breeding materials and screen more parental lines with multiple elite traits especially the key quantitative traits; (2) Determine the breeding objectives and crosses from the core germplasm group based on the inheritance of the receptor and donor parents with the aid of computer screening database. (3) Draw up the breeding pipeline and selection strategy based on breeding objectives and inheritance characters of the population with the aid of molecular markers (including QTLs, associated markers, and genes). (4) Select elite progeny of the crosses using molecular markers, precise trait evaluation techniques, and the elite traits in field under various environments. (5) Purify new varieties containing the target genotypes and phenotypes.

To excise the molecular breeding routine, the SGWG analyzed the QTL distribution of seed quality and disease resistance traits in sesame as

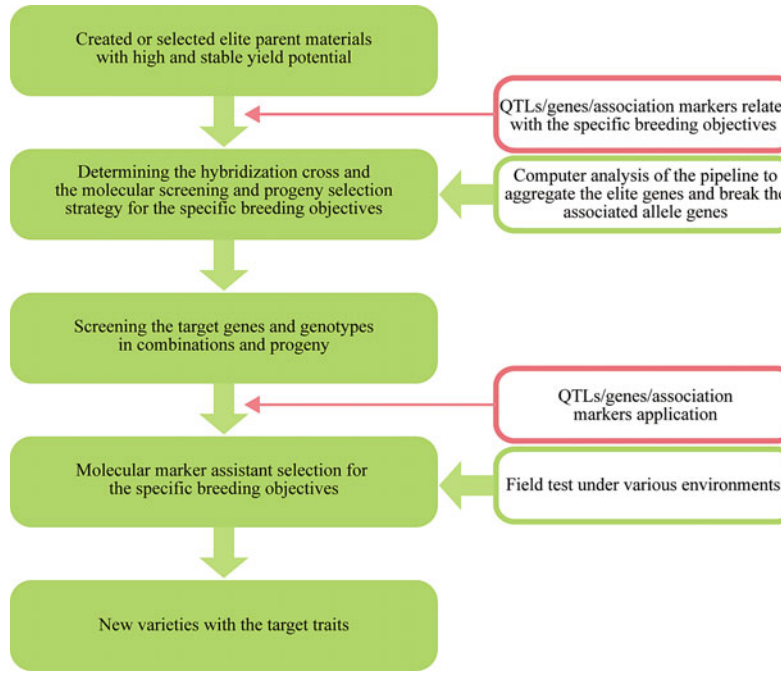


Fig. 19.1 Scheme of the molecular design breeding in sesame

an example. Recent genetic analysis indicated that a QTL of the 'thousand seed weight' trait was determined on the 63.5 cm of *SiChr.* 8 with the R^2 of 0.12, while a QTL of '*Fusarium* wilt disease resistance' trait was on 80.4 cm of *SiChr.* 8 with the R^2 of 0.18 using the ultra-dense SNP genetic map in sesame (Zhang et al. 2016) (unpublished data, Haiyang Zhang). The gene locus of seed weight trait is detected linked with the disease resistance alleles in sesame. Previous studied showed that only 10% sesame germplasm exhibit high resistance in *Fusarium* wilt nursery (Wang et al. 1999; Qiu et al. 2014).

Thus, in order to break the linkage of both traits and to breed a new variety with high seed weight and disease resistance level, 'Beijing Bawangbian' and 'XincaiXuankang' are chosen from the elite sesame lines library to construct a crossing population. 'Beijing Bawangbian' has the high thousand seed weight (TSW) of 4.53 g, but is highly susceptible to *Fusarium* wilt disease. 'XincaiXuankang' has high resistance to FOS pathogen but low TSW of 2.16 g. According to the heredity and the locus position, 10% possibility of the new lines with high TSW and disease resistance is predicted with the computer imitation, when screened using the molecular markers from the 400 lines of F_2 population, or 280 lines of F_3 population, or more than 200 lines of F_6 generations. Of course, to obtain the stable new lines, all the materials should be screened in the disease nursery field with various environmental conditions for further production application.

In a word, sesame Genome Project supplies the necessary and key genome information for molecular genetics research in sesame and will impel the genomics-assisted breeding. Comprehensive analysis of genome, epigenome, transcriptome, proteome and metabolome data of sesame will facilitate sesame breeding program in the future.

19.5 Conclusion

Genetics research develops rapidly in sesame in recent a few years. The initiation of the Sesame Genome Project impels the progresses of gene cloning, QTLs location, and GWAS association studies in sesame. The reference genome information facilitates the genome evolution and

phylogenetic analysis of the *Sesamum* species. We believe the accomplishment of the Sesame Genome Project will further improve the genetic and genomic research and the marker-assisted selection breeding in sesame. Correspondingly, more outstanding sesame varieties aggregating more elite traits would be bred through the efficient molecular breeding technology and meet with the industry development in near future.

References

Ashri A (2006) Sesame (*Sesamum indicum* L.). In: Signh RJ (ed) Genetic resources, chromosome engineering, and crop improvement. CRC Press, Boca Raton, FL, pp 231–280

Chen Z, Zhi Y, Yi M, Wang J, Liang X et al (1996) Transformation of engineered male sterile gene and establishment of transgenic plants in sesame (*Sesamum indicum* L.). Acta Agri Bor-Sin 11(4):33–38

Chyan CL, Lee TTT, Liu CP, Yang YC, Tzen JTC et al (2005) Cloning and expression of a seed-specific metallothionein-like protein from sesame. Biosci Biotechnol Biochem 69(12):2319–2325

Dossa K, Wei X, Li D, Fonceka D, Zhang Y et al (2016) Insight into the AP2/ERF transcription factor superfamily in sesame and expression profiling of DREB subfamily under drought stress. BMC Plant Biol 16 (1):171

Gupta PK, Kumar J, Mir RR, Kumar A (2010) Marker-assisted selection as a component of conventional plant breeding. Plant Breed Rev 33:145–217

Hata N, Hayashi Y, Okazawa A, Ono E, Satake H et al (2010) Comparison of sesamin contents and *cyp81q1* gene expressions in aboveground vegetative organs between two Japanese sesame (*sesamum indicum* L.) varieties differing in seed sesamin contents. Plant Sci 178(6):510–516

Hsiao ESL, Lin LJ, Li FY, Wang MMC, Liao MY et al (2006) Gene families encoding isoforms of two major sesame seed storage proteins, 11S globulin and 2S albumin. J Agric Food Chem 54:9544–9550

Jin UH, Lee JW, Chung YS, Lee JH, Yi YB et al (2001) Characterization and temporal expression of a ω-6 fatty acid desaturasec DNA from sesame (*Sesamum indicum* L.) seeds. Plant Sci 161:935–941

Kim MJ, Kim JK, Shin JS, Suh MC (2007) The SebHLH transcription factor mediates trans-activation of the SeFAD2 gene promoter through binding to E- and G-box elements. Plant Mol Biol 64:453–466

Kim MJ, Go YS, Lee SB, Kim YS, Shin JS et al (2010) Seed-expressed casein kinase I acts as a positive regulator of the SeFAD2 promoter via phosphorylation of the SebHLH transcription factor. Plant Mol Biol 73(s4–5):425–437

Kumpatla SP, Buyyarapu R, Abdurakhmonov IY, Mammadov JA (2012) Genomics-Assisted Plant breeding in the 21st Century: Technological advances and progress. In: Ibrokhim YA (Ed) Plant breeding, Intech, Rijeka, Croatia, pp 131–133

Lee TTT, Leu WM, Yang HH, Chen BCM, Tzen JTC (2006) Sesame oleosin and prepro-2S albumin expressed as a fusion polypeptide in transgenic rice were split, processed and separately assembled into oil bodies and protein bodies. J Cereal Sci 44(3):333–341

Miao H (2014) The Sesame Genome Project and sesame genome sequencing. In: Plant and animal genome XXII conference. 10–15th, January, San Diego, USA

Miao H, Zhang H (2016) The genome of *Sesamum indicu*m L. In: Plant and animal genome XXIV conference. Plant and Animal Genome, 9–13th, January, San Diego, USA

Qiu C, Zhang H, Chang S, Wei L, Miao H (2014) Laboratory detecting method for pathogenicity of *Fusarium oxysporum* Schl. f. sp. *sesami* isolates. Acta Phytopathol Sin 44(1):26–35

Varshney R, Song C, SaxenaR AS, Yu S et al (2013) Draft genome sequence of kabuli chickpea (*Cicer arietinum*): genetic structure and breeding constraints for crop improvement. Nat Biotechnol 31:240–246

Wang B, Dale ML, Kochman JK, Allen SJ, Obst NR (1999). Variations in soil population of *Fusarium oxysporum* f. sp. *vasinfectumas* influenced by fertiliser application and growth of different crops. Australas Plant Pathol 28(2):174–181

Wang L, Yu S, Tong C, Zhao Y, Liu Y et al (2014) Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. Genome Biol 15 (2):R39

Wang L, Xia Q, Zhang Y, Zhu X, Zhu X et al (2016) Updated sesame genome assembly and fine mapping of plant height and seed coat color QTLs using a new high-density genetic map. BMC Genomics 17(1):1–13

Wei X, Liu K, Zhang Y, Feng Q, Wang L et al (2015) Genetic discovery for oil production and quality in sesame. Nat Commun 6:8609

Yu J, Golicz AA, Lu K, Dossa K, Zhang Y et al (2019) Insight into the evolution and functional characteristics of the pan-genome assembly from sesame landraces and modern cultivars. Plant Biotechnol J 17(5):881–892

Yukawa Y, Takaiwa F, Yoshida N, Yamada K (1995) Structure and expression of a cDNA encoding steroyl-acyl carrier protein desaturase in sesame. Plant Cell Physiol 37:201–205

Zhang H, Miao H, Wei L, Li C, Zhao R et al (2013a) Genetic analysis and QTL mapping of seed coat color in sesame (*Sesamum indicum* L.). PLoS One 8(5): e63898

Zhang YX, Wang LH, Xin H, Li D, Ma C et al (2013b) Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. BMC Plant Biol 13(1):1–12

Zhang H, Miao HM, Li C, Wei LB, Duan YH et al (2016) Ultra-dense SNP genetic map construction and identification of *SiDt* gene controlling the determinate growth habit in *Sesamum indicum* L. Sci Rep 6:31556

Zhang H, Miao H, Ju M (2019) Potential for adaptation to climate change through genomic breeding in sesame. In: Kole C (ed) Genomic designing of climate-smart oilseed crops. Springer, Cham, pp 374–376