Compendium of Plant Genomes
Series Editor: Chittaranjan Kole

Chittaranjan Kole · Pablo Rabinowicz *Editors*

The Castor Bean Genome

Springer

Compendium of Plant Genomes

Series editor

Chittaranjan Kole, Raja Ramanna Fellow, Department of Atomic Energy, Government of India, ICAR-National Research Center on Plant Biotechnology, 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 70 plant species have been sequenced and genome sequences of several other plants are in the pipeline. Research publications on these genome initiatives are scattered on dedicated web sites and in journals with all too brief descriptions. The individual volumes elucidate the background history of the national and international genome initiatives; public and private partners involved; strategies and genomic resources and tools utilized; enumeration on the sequences and their assembly; repetitive sequences; gene annotation and genome duplication. In addition, synteny with other sequences, comparison of gene families and most importantly potential of the genome sequence information for gene pool characterization and genetic improvement of crop plants are described.

Interested in editing a volume on a crop or model plant? Please contact Dr. Kole, Series Editor, at ckole2012@gmail.com

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

Chittaranjan Kole · Pablo Rabinowicz
Editors

The Castor Bean Genome

Springer

Editors
Chittaranjan Kole
Raja Ramanna Fellow, Department of Atomic Energy, Government of India
ICAR-National Research Center on Plant Biotechnology
New Delhi, India

Pablo Rabinowicz
Rockville, MD, USA

ISSN 2199-4781 ISSN 2199-479X (electronic)
Compendium of Plant Genomes
ISBN 978-3-319-97279-4 ISBN 978-3-319-97280-0 (eBook)
https://doi.org/10.1007/978-3-319-97280-0

Library of Congress Control Number: 2018967744

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

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

Chittaranjan Kole

Preface to the Series

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

Physical mapping of genomes was the obvious consequence that facilitated 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 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 traveled much beyond biology and involves biophysics, biochemistry, and bioinformatics!

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

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

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

As I write this preface for the first volume of the new series "Compendium of Plant Genomes", a net search tells me that complete or nearly complete whole-genome sequencing of 45 crop plants, eight crop and model plants, eight model plants, 15 crop progenitors and relatives, and 3 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 is growing rapidly every day. However, the information is scattered in research articles and review papers in journals and dedicated Web pages of the consortia and databases. There is no compilation of plant genomes and the opportunity of using the information in sequence-assisted breeding or further genomic studies. This is the underlying rationale for starting this book series, with each volume dedicated to a particular plant.

Plant genome science has emerged as an important subject in academia, and the present compendium of plant genomes will be highly useful both to students and teaching faculties. Most importantly, research scientists involved in genomics research will have access to systematic deliberations on the plant genomes of their interest. Elucidation of plant genomes is 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, Dr. Christina Eckey and Dr. Jutta Lindenborn in particular, for all their constant and cordial support right from the inception of the idea.

I always had to set aside additional hours to edit books 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 to the Volume

Castor bean (*Ricinus communis*) is a member of the Euphorbiaceae or spurge family, which includes other economically and scientifically important species, such as the staple crop cassava, the ornamental poinsettia, the parasitic *Rafflesia* that produces the largest flower among angiosperms, and the industrially important rubber tree, among others. Despite its name, castor bean is not legume but it is an important oilseed crop because of the high proportion of the unusual hydroxylated fatty acid ricinoleate found in its oil. The high unusual fatty acid content confers castor oil unique properties that make it extremely valuable for multiple industrial and chemical applications.

The castor bean genome sequence published in 2010 was the first to be completed within the Euphorbiaceae family. Therefore, it provided not only in-depth insights into the biology of this important crop, but also a better understanding of the Euphorbiaceae family in general, which is an important and, until then, understudied plant group. However, the main motivation to initiate the castor bean genome sequencing project was safety concerns. In addition to oil, the castor bean seeds accumulate large amounts of ricin, a highly toxic ribosome inactivation protein that has been reportedly used in the 1970s to assassinate a Bulgarian broadcaster in London using an injection device disguised as an umbrella. Allegedly, the attack deployed 0.28 mm^3 of ricin into the victim's bloodstream, and although no traces of any poison could be detected, ricin was assumed to be the culprit because no other known substance could kill a man in such small amounts.

Even though ricin is one of the deadliest toxins known when inhaled or injected into the bloodstream, it is less toxic if ingested due to proteolytic inactivation in the digestive tract. Therefore, it would be very difficult to use ricin as a bioweapon and no such incidents have been reported. Nevertheless, the presence of ricin as well as other toxic agglutinins and allergens in its seed poses security challenges for castor bean cultivation, industrial processing, and use as animal feed. The castor bean genome sequence has uncovered previously unknown details about the ricin gene family and has opened the door for comparative studies using natural diversity as well as, eventually, genome-wide association studies (GWAS) to develop varieties with low ricin content. Such cultivars would reduce health risks for workers involved in castor bean cultivation and related industries and would expand the potential use of castor bean processing waste as animal feed.

The improvements of castor bean that its genome sequence could catalyze are not limited to reducing toxicity. Developing new cultivars with increased and/or altered castor oil production to achieve high yields of tailored fatty

acids can be facilitated by the availability of the castor bean genome sequence, which has increased our knowledge of castor bean lipid metabolic pathways as well as the regulatory networks that control them. New castor bean lines with improved oil properties combined with low toxicity would revitalize the castor oil industry that has been negatively affected by the biosafety concerns posed by ricin, particularly in the USA.

Furthermore, with the advent of plant genome editing technologies, irreversibly eliminating all members of the ricin and allergen gene families is becoming a real short-term possibility, and high-throughput genome resequencing technologies can be easily used to analyze the genetic makeup of natural diversity or newly engineered varieties.

This volume intends to cover multiple aspects of castor bean's biology, from anatomical descriptions to genetic engineering. Different chapters highlight how the availability of the genome sequence has enabled new research in this important oilseed crop, including the development of genetic markers and maps, mining for relevant gene families such as disease resistance genes, oil biosynthesis and metabolomics, epigenetics, and the generation of transgenic plants for different purposes. While each chapter is meant to be an independent read, we hope that the diverse topics addressed in this book will be of interest to researchers within the specific field as well as in other areas of plant biology.

We are thankful to all the authors that contributed their work to put together this volume.

New Delhi, India — Chittaranjan Kole
Rockville, USA — Pablo Rabinowicz

Contents

Contributors

Kwadwo Gyapong Agyenim-Boateng Agricultural College, Guangdong Ocean University, Zhanjiang, Guangdong, China

Ch. Anil Kumar ICAR-Indian Institute of Oilseeds Research, Hyderabad, India;
Department of Genetics, Osmania University, Hyderabad, India

Mohd. Ashraf Ashfaq ICAR-Indian Institute of Oilseeds Research, Hyderabad, India;
ICAR-Indian Institute of Rice Research, Hyderabad, India;
Department of Plant Sciences, University of Hyderabad, Hyderabad, India

Rajib Bandopadhyay Department of Botany, UGC-CAS, The University of Burdwan, Burdwan, West Bengal, India

Danilo L. J. Batista Metabolomics Research Group, Instituto de Química, Universidade Federal da Bahia, Salvador, Brazil

Valdinei C. Brito Metabolomics Research Group, Instituto de Química, Universidade Federal da Bahia, Salvador, Brazil

Cristiane D. de Brito Laboratório de Bioquímica, Biotecnologia E Bioprodutos, Departamento de Bioquímica E Biofísica, Universidade Federal da Bahia, Salvador, Brazil

Gisele A. B. Canuto Metabolomics Research Group, Instituto de Química, Universidade Federal da Bahia, Salvador, Brazil

Renato Delmondez de Castro Laboratório de Bioquímica, Biotecnologia E Bioprodutos, Departamento de Bioquímica E Biofísica, Universidade Federal da Bahia, Salvador, Brazil

Agnes P. Chan J. Craig Venter Institute, Rockville, MD, USA

Rajinder Singh Chauhan Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan, India;
Department of Biotechnology, Bennett University, Greater Noida, Uttar Pradesh, India

Grace Q. Chen Agricultural Research Service, US Department of Agriculture, Western Regional Research Center, Albany, CA, USA

V. Dinesh Kumar ICAR-Indian Institute of Oilseeds Research, Hyderabad, India

Bhramar Dutta Department of Botany, UGC-CAS, The University of Burdwan, Burdwan, West Bengal, India

Luzimar Gonzaga Fernandez Laboratório de Bioquímica, Biotecnologia E Bioprodutos, Departamento de Bioquímica E Biofísica, Universidade Federal da Bahia, Salvador, Brazil

Federico García Maroto Grupo de Investigación "Biotecnología de Productos Naturales" (BIO-279), CEIA3, BITAL, Universidad de Almería, Almería, Spain

Henk W. M. Hilhorst Wageningen Seed Lab, Laboratory of Plant Physiology, Wageningen University, Wageningen, The Netherlands

Kumiko Johnson Agricultural Research Service, US Department of Agriculture, Western Regional Research Center, Albany, CA, USA

Prathap Reddy Kallamadi Crop Improvement Section, ICAR-Indian Institute of Oilseeds Research, Hyderabad, India

C. Lavanya ICAR-Indian Institute of Oilseeds Research, Hyderabad, Telangana, India

Wilco Ligterink Wageningen Seed Lab, Laboratory of Plant Physiology, Wageningen University, Wageningen, The Netherlands

Aizhong Liu Department of Economic Plants and Biotechnology, and Yunnan Key Laboratory for Wild Plant Resource, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China;
College of Forestry, Southwest Forestry University, Kunming, China

Diego López Alonso Grupo de Investigación "Biotecnología de Productos Naturales" (BIO-279), CEIA3, BITAL, Universidad de Almería, Almería, Spain

Marta B. Loureiro Laboratório de Bioquímica, Biotecnologia E Bioprodutos, Departamento de Bioquímica E Biofísica, Universidade Federal da Bahia, Salvador, Brazil

Jiannong Lu Agricultural College, Guangdong Ocean University, Zhanjiang, Guangdong, China

Eva Morale Agricultural Research Service, US Department of Agriculture, Western Regional Research Center, Albany, CA, USA

Bhimasen Naik Regional Research and Technology Transfer Station, Orissa University of Agriculture and Technology, Sambalpur, Odisha, India

Paulo R. Ribeiro Laboratório de Bioquímica, Biotecnologia E Bioprodutos, Metabolomics Research Group, Departamento de Bioquímica E Biofísica, Instituto de Química, Universidade Federal da Bahia, Salvador, Brazil

Velu Mani Selvaraj ICAR-Indian Institute of Oilseeds Research, Hyderabad, India;
Department of Plant Sciences, University of Hyderabad, Hyderabad, India

P. Soma Sekhar Reddy ICAR-Indian Institute of Oilseeds Research, Hyderabad, India;
Department of Genetics, Osmania University, Hyderabad, India

Archit Sood Biotechnology Division, CSIR-Institute of Himalayan Bioresource Technology, Palampur, Himachal Pradesh, India

Mulpuri Sujatha Crop Improvement Section, ICAR-Indian Institute of Oilseeds Research, Hyderabad, Telangana, India

Daniele Takahashi Laboratório de Bioquímica, Biotecnologia E Bioprodutos, Departamento de Bioquímica E Biofísica, Universidade Federal da Bahia, Salvador, Brazil

Meilian Tan Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute (OCRI) of Chinese Academy of Agricultural Sciences, Wuhan, People's Republic of China

Muddanuru Tarakeswari ICAR-Indian Institute of Oilseeds Research, Hyderabad, Telangana, India

Rukam Singh Tomar Department of Biotechnology, Junagadh Agricultural University, Junagadh, Gujarat, India

A. Vishnuvardhan Reddy ICAR-Indian Institute of Oilseeds Research, Hyderabad, Telangana, India

Lei Wang Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute (OCRI) of Chinese Academy of Agricultural Sciences, Wuhan, People's Republic of China

Wei Xu Department of Economic Plants and Biotechnology, and Yunnan Key Laboratory for Wild Plant Resource, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China

Xingchu Yan Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute (OCRI) of Chinese Academy of Agricultural Sciences, Wuhan, People's Republic of China

Xuegui Yin Agricultural College, Guangdong Ocean University, Zhanjiang, Guangdong, China

Zhi Zou Key Laboratory of Biology and Genetic Resources of Tropical Crops, Ministry of Agriculture and Rural Affairs, Institute of Tropical Biosciences and Biotechnology, Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan, People's Republic of China

Abbreviations

10KP	10,000 Plants Genomes Project
1KP	1000 Plants Initiative
3C	Chromosome conformation capture
4mC	4-methylcytosine
5hmC	5-hydroxylmethylcytosine
5mC	5-methylcytosine
6mA	6-methyladenine
AFLP	Amplified fragment length polymorphism
ALP	Aleurain-like protease
AMOVA	Analysis of molecular variance
AP2	*APETALA2* (gene)
AQP	Aquaporin
Ar/R	Aromatic/arginine
ASTM	The American Society of Testing and Materials
ATP	Adenosine triphosphate
BA	Benzyladenine; N^6-benzylamino purine
BAC	Bacterial artificial chromosome
Bar	Bialaphos resistance gene
BFA	Brefeldin A
BRH	Best reciprocal hit
BS-seq	Bisulfite sequencing
Bt	*Bacillus thuringiensis*
bZIP	Basic leucine zipper
C	Carboxyl terminus
CaM	Calmodulin
CaMV	Cauliflower mosaic virus
Cas9	Native Cas9 nuclease
CB-1A	Castor bean allergen
CC	Coiled coil
CCA	Canonical correlation analysis
CCV	Clathrin-coated vesicle
CE	Capillary electrophoresis
CEP	Cysteine endopeptidase

Chr	Chromosome
CMA	Chromomycin
CMT3	Chromomethylase 3
CNL	Coiled-coil-NBS-LRR
CoA	Coenzyme A
CPT	Choline phosphotransferase
CRISPR	Clustered regularly interspaced short palindromic repeats
cry1AcF	Delta-endotoxin of *Bt* gene (1AcF)
cry1EC	Delta-endotoxin of *Bt* gene (1EC)
CTB	Cathepsin B-like
DAP	Days after pollination
DAPI	4,6,-diaminido-2-phenylindole
DAS	Days after sowing
DEGs	Differentially expressed genes
DGAT	Diacylglycerol actetyltransferase/acyltransferase
DGE	Digital gene expression
DGTA	Diacylglycerol: diacylglycerol transacylase
DI	Diversity index
DME	DNA glycosylase DEMETER
DML	DEMETER-like (protein)
DNB	DNA nanoball
DNRs	Dinucleotide repeats
DOE	Department of Energy (USA)
Dof	DNA binding with one finger
DRM1/2	Domains rearranged methyltransferase 1 and 2
dsRNA	Double-stranded RNA
DUOX	Dual oxidase
ECS	Endocytosis cell signaling domain
EI	Electron ionization
ELISA	Enzyme-linked immune sorbent assay
eLRR	Extracellular LRR
EMBRAPA	Brazilian Agricultural Research Company
EMR	Effective multiplexing ratio
ER	Endoplasmic reticulum
ERAD	ER-associated degradation
EREBP	Ethylene-responsive element binding protein
ERF	Ethylene-responsive (element binding) factor
EST	Expressed sequence tag
FA	Fatty acid
FAD2	Fatty acid desaturase 2
FAE	Fatty acid elongase
FAH	Fatty acid hydroxylase
FAH12	Fatty acid hydroxylase 12
FISH	Fluorescence in situ hybridization
FPKM	Fragments per kilobase of exon per million fragments mapped

G3P	Glycerol-3-phosphate
GA	Gibberellic acid
GABA	Gamma-aminobutyric acid
GAD	Glutamate decarboxylase
GC	Gas chromatography
GC	Guanine/cytosine
GC-TOF-MS	GC coupled to a quadrupole time-of-flight MS
GFP	Green fluorescent protein
GISH	Genomic in situ hybridization
GO	Gene ontology
GPAT	Glycerol-3-phosphate acyltransferase
GRAS	GAI, RGA, and SCR
GSC	Genome Sequencing Center (USA)
GUS	β-glucuronidase
H3K27me3	Histone H3 lysine 27 trimethylation
H3K9me2	Histone H3 lysine 9 dimethylation
HE	Expected value of heterozygosity
HNRs	Hexanucleotide repeats
HO	Observed value of heterozygosity
HPLC	High-performance liquid chromatography
HPT	Hygromycin phosphotransferase
HTRs	Heptanucleotide repeats
ICAR	Indian Council of Agricultural Research
IGS	University of Maryland Institute of Genome Science (USA)
ihpRNA	Intron hairpin RNA
IIOR	Indian Institute of Oilseeds Research
ISF	Interspersed staminate flowers
ISS	International Space Station
ISSR	Inter-simple sequence repeat
JCVI	J. Craig Venter Institute (USA)
JGI	Joint Genome Institute (USA)
KAS	3-keto-acyl-ACP synthase
LACS	Long-chain acyl-CoA synthetase
LC	Liquid chromatography
LC-MS	Liquid chromatography-MS
LC-MS/MS	LC-tandem MS
LEC	Leafy cotyledon
LG	Linkage group
Lhc	Light-harvesting chlorophyll a/b-binding

LOD	Logarithm of odds
LPA	Lysophosphatidic acid
LPAT	Lysophosphatidate acyltransferase
LPC	Lysophosphatidylcholine
LPCAT	Lysophosphatidylcholine acyltransferase
LRR	Leucine-rich repeat
LSC	Large single-copy
M	Monoecious
MALDI-TOF	Matrix-assisted laser desorption ionization time-of-flight
MAS	Marker-assisted selection
MASP	Methylation-sensitive amplification polymorphism
Mb	Megabase
MBOAT	Membrane-bound O-acyltransferase
Mbp	Megabase pair
Mcr	Modified cytosine restriction system
MDA	Malondialdehyde
MEG	Maternally expressed genes
MET1	Methyltransferase 1
MF	Methylation filtration
MHC	Major histocompatibility complex
MI	Marker index
MIP	Major intrinsic protein
MRM-MS	Multiple reaction monitoring mass spectrometry
mRNA	Messenger RNA
MS	Mass spectrometry
MW	Molecular weight
Mya	Million years ago
N	Amino terminus
NAA	Naphthaleneacetic acid
NAM	NAC transcription factor
NBS	Nucleotide-binding site
NBS-LRR-CC	NBS-LRR-coiled-coil
NCBI	National Center for Biotechnology Information (USA)
ncRNA	Non-coding RNA
NES-type	Nebraska and sex-revertant pistillate mechanism
NIP	NOD26-like intrinsic protein
NIST	National Institute for Standards and Technology (US Department of Commerce)
NLS	Nuclear localization signal
NMR	Nuclear magnetic resonance
NOR	Nuclear organization region
NPA	Asparagine–proline–alanine
*npt*II	Neomycin phosphotransferase II gene

N-terminal	Amino-terminal
N-type	Nebraska-type pistillate mechanism
OA	Oleic acid
OG	Orthologous group
OGG	Open Green Genomes Initiative (DOE, USA)
OLE	Oleosin
Ole-CoA	Oleoyl-CoA
Ole-PC	Sn-2-oleoyl-phosphatidylcholine
ONT	Oxford Nanopore Technology
P	Pistillate
PAGE	Polyacrylamide gel electrophoresis
PAH	Phosphatidic acid hydrolase
PASA	Program to Assemble Spliced Alignments
PC	Phosphatidylcholine
PCA	Principal component analysis
PCR	Polymerase chain reaction
PDAT	Phospholipid:diacylglycerol acyltransferase
PEG	Paternally expressed genes
PEST	Amino acid domain
PIC	Polymorphic information content
PIP	Plasma membrane intrinsic protein
PLA2	Phospholipase A2
PLC	Phospholipase C
PLCP	Papain-like cysteine protease
P-loop	Phosphate-binding-loop
PLS	Partial least squares
PNRs	Pentanucleotide repeats
PRC2	Polycomb group proteins
PSV	Protein storage vacuole
PTGS	Post-transcriptional gene silencing
QTL	Quantitative trait locus
QTLs	Quantitative trait loci
R	Resistance
RA	Ricinoleic acid
RAPD	Random(ly) amplified polymorphic DNA
RCA	*Ricinus communis* agglutinin
RcFBPase	Fructose-1,6-bisphosphatase
RcPEPCK	Phosphoenolpyruvate carboxykinase
RcPGM	Phosphoglucomutase
RcSS	Starch synthase
RD19 (21)	Responsive to dehydration 19 (21)
RdDM	RNA-directed DNA methylation
RFLP	Restriction fragment length polymorphism
RGA	Resistance gene analog
Ric-CoA	Ricinoleoyl-CoA
RIP	Ribosome-inactivating protein
RMAPD	Random microsatellite amplified polymorphic DNA

RNAi	RNA interference
RNA-seq	RNA sequencing
RNN	Recurrent neural network
ROS1	Repressor of Silencing 1
RP	Resolving power
Rpg1	Barley stem rust resistance protein
RPM1	Resistance to *Pseudomonas syringae* 3
RPP5	Recognition of *Peronospora parasitica* 5
RPS2	Resistance to *Pseudomonas syringae* 2
RPW8	Resistance to powdery mildew 8
RRS1R	Resistance to *Ralstonia solanacearum* 1
SAD	Stearoyl acyl carrier protein desaturases
SAG12	Senescence-associated gene 12
SbNHX1	*Salicornia brachiata* reverse transporter protein gene
SBP	SQUAMOSA-promoter-binding protein
SCoT	Start codon targeted
SD	Standard deviation
SDS	Sodium dedocyl sulfate
S-genes	Susceptibility genes
SIP	Small basic intrinsic protein
siRNA	Small interfering RNA
SLAF	Specific length amplified fragment
SMRT	Single-molecule real-time
SNP	Single-nucleotide polymorphism
SRAP	Sequence-related amplified polymorphism
SSC	Small single-copy
SSR	Simple sequence repeat
ssVSS	Sequence-specific vacuolar sorting signal
S-type	Sex-revertant pistillate mechanism
TAG	Triacylglycerol
TDZ	Thidiazuron
TE	Transposable element
TFA	Total fatty acids
TFL-like	Terminal flower-like
TIGR	The Institute for Genomic Research (USA)
TILLING	Targeting induced local lesions in genomes
TIP	Tonoplast intrinsic protein
TIR	TOLL/interleukin-1 receptor
TMS	Trimethylsilyl
TNL	TIR-NBS-LRR
TNRs	Trinucleotide repeats
TRAP	Target region amplification polymorphism
TrD	Transmembrane domain
TTRs	Tetranucleotide repeats
UFA	Unusual fatty acid
USDA	United States Department of Agriculture
VIP	Vegetative insecticidal *Bt* proteins

VPE	Vacuolar processing enzyme
WGD	Whole-genome duplication
WGS	Whole-genome shotgun sequencing
WRKY	WRKY domain/transcription factor
WUS	WUSCHEL gene
Xa21	Receptor kinase-like protein
XBCP3	Xylem bark cysteine peptidase 3
XCP	Xylem cysteine peptidase
XIP	X intrinsic protein

Botanical Descriptions of Castor Bean

1

Bhimasen Naik

Abstract

Castor bean, *Ricinus communis* L., belongs to the spurge family (Euphorbiaceae). Despite its name, the seed is really not a true bean and it is not related to the bean or legume family Fabaceae. It is an oilseed crop cultivated mainly in India, Mozambique, Brazil, and China; and believed to have polyphyletic origin with four centers of diversity. The plant is an annual herb, or a perennial shrub or small tree. Blooms are found on the stem and certain other parts of the castor bean plant. The inflorescence is an erect and terminal panicle of cymes (panicled cymes). The flowers are usually unisexual and monoecious. The staminate and pistillate flowers are borne on the same inflorescence. Castor bean has a mixed mating system generating both selfed and cross-fertilized offspring. It is basically a long-day plant, but is adaptable, with less yields, to a wide range of photoperiods. The fruit is botanically a "schizocarpic capsule" or regma. The seed is ovoid, tick-like, caruncu-late, albuminous, poisonous, and allergenic. The germination is epigeal. The oil extracted from the seeds is non-drying in nature with a lot of uses in medicine, cosmetics, biodiesel, and other industries. The detoxified castor bean meal and husk are used as animal feed. The castor bean meal is also an organic manure. The active poison in castor bean seeds is ricin, a very deadly protein called lectin. Ricin is found in the meal or cake after the oil has been extracted. It is not carried over into the oil if it is properly extracted, but remains in the meal. The leaves contain ricin, but in much smaller quantities than in the seeds.

1.1 Introduction

Castor bean (syn. castorbean, castor, castor-oil-plant), *Ricinus communis* L. ($2n = 20$, $X = 10$), is a species of flowering plant in the spurge family, Euphorbiaceae. It is an oilseed crop cultivated mainly in India, Mozambique, Brazil, and China (FAOSTAT 2014). It is interesting to trace the origin of the name "castor". Castor is the generic name of the North American beaver (*Castor canadensis*) and one of the brightest double stars in the constellation Gemini. In Greek and Roman legend, castor was one of the twin sons of Jupiter and Leda. According to Weiss (1971), the name "castor" has nothing to do with beavers, luminous stars, or offspring of Greek and Roman Gods. Castor was apparently coined by English traders who confused it with the oil of another shrub,

B. Naik (✉)
Regional Research and Technology Transfer Station, Orissa University of Agriculture and Technology, Chiplima, Sambalpur 768 025, Odisha, India
e-mail: bsnaikouat1@gmail.com

C. Kole and P. Rabinowicz (eds.), *The Castor Bean Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97280-0_1

Vitex agnus-castus, which the Spanish and Portuguese in Jamaica called "agno-casto". Although it is commonly known as castor bean plant, the seed is really not a true bean and it is not related to the bean or legume family, Fabaceae. There are many other examples of "beans" that are botanically not beans, such as Mexican jumping beans and coffee beans. It has been proposed that the term "bean" should be discontinued in favor of castor plant and castor seed (Weiss 2000). Avoiding the use of the term bean is really important because the seed and whole plant are very poisonous and should not be eaten. To call simply "castor" is logical.

1.2 Origin and Distribution

1.2.1 Origin

There are various opinions regarding the center of origin of castor bean. Hooker (1890) is of opinion that though castor bean is cultivated throughout India, it is indigenous to Africa. De Candolle (1860) also has a similar view. But Fluckiger and Hanbury (*fide* Watt 1892) affirm that it is native to India. They hold this view mainly on the basis of knowledge of the medicinal uses of this plant as found in Sanskrit literature. Bentley and Trimen (*fide* Watt 1892) also believe that castor bean is native to India. Watt (1892) in his extensive tour of India found some evidence only at the foot of the Himalayas to show that castor bean is native of India. He, therefore, believed that it might have originated in India as well as in Africa. Hindus have known castor oil from very ancient times. This oil has been mentioned in *Susruta Ayurveda*, one of the oldest works on Ayurveda. It is, therefore, possible that castor bean had originated in India or Africa.

Castor bean is also believed to have polyphyletic origin with four centers of diversity, viz, (i) Ethiopian-East African, (ii) North–West and South–West Asia and Arabian Peninsula, (iii) Subcontinent of India, and (iv) China. However, Ethiopian-East African region is considered to be the most probable site of origin (Moshkin 1986).

1.2.2 Distribution

Wild varieties of castor bean grow in East and North Africa, and the Near and Middle East. It was cultivated in ancient Egypt as long ago as 4000 BC. It was taken at an early date to India and beyond, and was recorded in China in the Tang period (618–906 AD). It was introduced into the New World shortly after Columbus (Purseglove 1968). According to Weibel (1948), the oldest record of castor bean presence in the USA was in 1818 in Illinois. The castor bean plant is now naturalized in many tropical and subtropical countries.

1.3 Taxonomy

1.3.1 Taxonomy Tree

According to the phylogenetic system of classification of Hutchinson (1959), castor bean is classified as:

Phylum—Angiospermae
Subphylum—Dicotyledones
Division—Lignosae
Order—Euphorbiales
Family—Euphorbiaceae
Group—Platylobeae
Subfamily—Crotonoideae
Tribe—Acalypheae
Subtribe—Ricininae
Genus—*Ricinus* L.
Species—*communis*
Binomial name—*Ricinus communis* L.

1.3.2 Classification

According to Hutchinson, Euphorbiales is the thirty-fifth order of the phylum Angiospermae, subphylum Dicotyledones and division Lignosae. The order consists of only one family, i.e., Euphorbiaceae (spurge family), which is the fourth largest Angiosperm family, and it includes 218 genera and 6745 species distributed

worldwide (Pandey 1976; Shukla and Misra 1982; http://www.cabi.org/isc/datasheet/47618).

1.3.2.1 Salient Features of Euphorbiaceae

Trees or shrubs, a few herbs. Milky latex invariably present. Leaves generally simple, stipulate, alternate. Inflorescence variable. Flower unisexual, hypogynous, and regular. Perianth 1–2 whorls, rarely absent. Stamen 1–∞, free or connate. Ovary trilocular, placentation axile, with 1–2 ovules in each loculus, style as many as carpels, stigma bifid. Fruit schizocarpic, regma. Seeds albuminous (Shukla and Misra 1982).

1.3.2.2 Divisions of the Family

Pax and Hofmann divided Euphorbiaceae into two groups (Shukla and Misra 1982).

Group A—*Platylobeae*. Cotyledons broader than radicle.

Subfamily I. *Phyllanthoideae*. Ovule two in each ovary cell. Latex vessel absent.
Tribe 1. *Bridelieae*. Sepals in male flower valvate, e.g., *Bridelia*, *Cleistanthus,* etc.
Tribe 2. *Phyllantheae*. Sepals in male flower imbricate, e.g., *Phyllanthus*, *Bischofia.*
Subfamily II.*Crotonoideae*. Ovule one in each ovary cell. Latex vessel present or absent.
Tribe1. *Chrozophoreae*. Petals in male flowers present. Sepals of the same valvate. Stamens in 1–3 whorls, e.g., *Aleurites, Chrozophora*.
Tribe 2. *Crotoneae*. Sepals almost open, valvate, or imbricate. Petals in male flowers present. Filaments bent in bud, e.g., *Croton.*
Tribe 3. *Cluytieae*. Flowers with petals, sepals of the flower imbricate. Inflorescence always cymose, e.g., *Jatropha*, *Cluytia*.
Tribe 4. *Foannesieae*. Leaves digitately compound. Petals absent. Calyx copular, shortly toothed, e.g., *Hevea*.
Tribe 5. *Acalypheae*. Sepals in male flowers valvate. Petals absent. Inflorescence racemose, e.g., *Acalypha, Ricinus*.
Tribe 6. *Manihoteae*. Calyx of male flower petaloid, gamosepalous, e.g., *Manihot*.
Tribe 7. *Gelonieae*. Petals absent. Sepals in male flower imbricate. Male flowers clustered or racemose, e.g., *Gelonium.*
Tribe 8. *Hippomaneae*. Calyx in the male very much reduced, segments open during flowering, e.g., *Sapium, Hippomane*.
Tribe 9. *Dalechampeae*. Inflorescence cymes of male and female flowers surrounded by leafy involucre. Stigma simple, e.g., *Dalechampia.*
Tribe 10. *Euphorbeae*. Inflorescence cyathium. Flowers very much reduced, naked, e.g.,. *Euphorbia, Pedilanthus.*

Group B—*Stenolobeae*. Cotyledons narrow and as broad as radicle.

Subfamily I. *Porantheroideae*. Ovules two in each loculus, e.g., *Poranthera.*
Subfamily II. *Ricinocarpoideae*. Ovules one in each loculus, e.g., *Ricinocarpus.*

1.3.2.3 The Genus *Ricinus* L

The genus *Ricinus* L. is considered to be monotypic. Previously described species within the genus *Ricinus* have been transferred to other genera or grouped within *R. communis*. The scientific name for the castor bean plant, *Ricinus communis* L., has a logical derivation. *Communis* means "common" in Latin, and castor bean plants were already commonly naturalized in many parts of the world when the eighteenth century Swedish naturalist Carolus Linnaeus was assigning scientific first and last names to plants and animals over 200 years ago. *Ricinus* is the Latin word for tick and the specific epithet for Mediterranean sheep tick (*Ixodes ricinus*). Apparently, Linnaeus thought that the seeds looked like ticks, particularly large ticks engorged with blood. The mottled body of certain ticks superficially resembles the castor bean seed (especially when the tick is engorged with blood), and the tick's head resembles the caruncle of a castor bean seed (http://waynesword.palomar.edu/plmar99.htm, http://www.cabi.org/isc/datasheet/47618).

1.4 Morphology

1.4.1 Habit

The plant is an annual herb or a perennial shrub, 1–7 m high (Fig. 1.1), glabrous. In the wild, it reaches the size of a small tree (Purseglove 1968). The plant height is highly influenced by the environment. In poor soil and scarcity of moisture, the plant attains a height of 30–90 cm; but the same cultivar, if sown in fertile soil with good rainfall, may attain a height of 3–4 m. There are types that are comparatively dwarfed even under favorable conditions. They usually attain a height of only 60–90 cm (Kulkarni and Ramanamurthy 1977).

1.4.2 Root

The plant has well-developed tap root with prominent laterals which produce a surface mat of feeding roots (Purseglove 1968). The tap root looks like an extension of the stem below the soil. It can reach up to 5 m underground in extremely poor soil. The secondary roots are restricted to the upper 75 cm of the soil. They mostly travel parallel to the ground, with a slight bend downward and grow to about 90–120 cm. The tertiary roots are not very long, hardly 30–45 cm. Root hairs have not been noticed in castor bean. The short duration types show high rate of root development during the early period, while medium- and long-duration types exhibit later. The long-duration types show longer lateral roots with greater penetration as compared to short- and medium-duration types (Kulkarni and Ramanamurthy 1977).

1.4.3 Stem

The stem is erect, cylindrical, branched, often brittle, glabrous, glaucous, green, or reddish, with waxy deposits (blooms), solid (pithy) becoming hollow with age, with well-marked nodes and prominent leaf scars, internodes shortest at base of the stem. A single stem is first produced which terminates in inflorescence at 6–10th node in dwarf early cultivars, at 8–16th node in later-maturing cultivars, and at 40th or more node in tall and wild plants. As the panicle develops, 2–3 sympodial branches grow out, one from each node immediately below it; they end in inflorescences and one or more sympodial branches grow out from the nodes immediately below them and the process continues. Thus, the development along each axis is sequential and a plant will have inflorescences at various stages of development. Degree of branching varies considerably (Purseglove 1968; Pandey 1976; Sundararaj and Thulasidas 1976; Shukla and Misra 1982). The branching character is very much influenced by environmental factors such as type

Fig. 1.1 Castor plant with different heights. *Courtesy* Milani and Nobrega (2013)

of soil, fertility of soil, moisture, and spacing. (Kulkarni and Ramanamurthy 1977).

1.4.3.1 Color of the Stem

The color of the stem is variable. White (1918) classified the stem color into five categories, viz, (i) bright green, (ii) green with reddish-bluish on the sunny side, (iii) carmine or rose-red, (iv) mahogany red, and (v) purple (dark red). Seshadri and Varisai (1951) reported four distinct stem colors, viz, mahogany, red, red-bluish and green. In the absence of a unified code, different workers recognized different grades of stem color.

The waxy bloom gives red or green stem a bluish appearance in the field. The stem color may be green, red or purple, and every gradation of the color may be noticed. It generally turns into gray-like color at the base when the castor bean plant is old. The presence of plastids in the stem at juvenile stage gives opportunity for additional photosynthetic activity (Salihu et al. 2014).

1.4.3.2 Bloom on the Stem

The blooms (waxy coating) are found on the stem and certain other parts of the castor bean plant. The color of the bloom is ashy white. The types of bloom found on castor bean are described under the following categories:

No bloom	With no bloom on any part of the plant.
Single-bloom	With bloom only on stem, but not on leaves or fruits.
Double-bloom	With bloom only on stem, fruits, and on the lower (dorsal) surface of leaves, not on the upper (ventral) surface of leaves.
Triple-bloom	With bloom on every part such as stem, petiole, both upper and lower surfaces of leaves, and fruits.

Narain (1952) described different grades of bloom in castor bean.

Most of the cultivars of castor bean in India comes under the double-bloom and triple-bloom groups (Kulkarni and Ramanamurthy 1977).

Bloom is a natural protection to the plant from extremes of weather and from some insect pests. Castor bean raised in the winter (*rabi*) season (October–May) showed that plant without bloom suffered more from cold than plant with bloom and also showed that the former suffered a more severe attack of jassids than the latter (Kulkarni and Ramanamurthy 1977). Cultivars with a heavy bloom are more resistant to jassids and leaf hoppers than bloomless types (Seshadri and Seshu 1956; Brar et al. 1977).

1.4.4 Leaf

The leaves are petiolate, stipulate, dorsiventral (bilateral, bifacial), peltate, simple and palmately partite, alternate and 2/5 in phyllotaxy, and glossy. The size of the leaf varies in different cultivars. Some cultivars are characterized by large leaves whereas others bear only small leaves. The length of the leaf ranges from about 15 to 45 cm. The petiole is fistular, 8–50 cm long, round, pale green or reddish, and with two nectiferous glands at junction to lamina, two glands on either side at leaf base and one or more glands on upper surface toward leaf base. The stipules are 1–3 cm long, united, broad, prominent, sheathing the bud, and deciduous. The leaves are spirally arranged except two opposite leaves at the node immediately above cotyledons. The leaf lamina (blade) is orbicular, 10–75 cm in diameter, and with 5–11 partite lobes for about half length. The lobes (segments) are ovate or lanceolate; lobe-apex acuminate; margin serrate; venation reticulate, multicostate, divergent, prominent veins on the lower surface; dark green or reddish above, paler green beneath. The leaf is usually green, but it is associated with the color of the stem. In case of green stem, the leaf lamina and mid-ribs are all green. In case of red stem, the young leaves have red tinge which becomes green when the leaf is fully developed; but the mid-rib maintains the reddish tinge. The leaf color varies from light green to dark red depending on the level of anthocyanin pigmentation present. In some castor bean varieties, the leaves start off as dark-reddish-purple or bronze when young but gradually changing to a dark green, sometimes with a reddish tinge as they

mature. The leaves of some others are green practically from the start, whereas in yet others a pigment masks the green color of all the chlorophyll-bearing parts such as leaves, stems and young fruits, so that they remain a dramatic purple-to-reddish-brown throughout the life of the plant (Purseglove 1968; Sundararaj and Thulasidas 1976; Kulkarni and Ramanamurthy 1977; Salihu et al. 2014).

1.4.5 Inflorescence

Usually, the main shoot and the branches terminate in inflorescences. The inflorescence is erect, 10–40 cm long, and terminal panicle of cymes (panicled cymes). Under adverse soil and rainfall conditions, inflorescences often emerge when plants are hardly 20–30 cm tall, while the same cultivar under optimum soil and weather conditions gives out main panicle at a height of 180–220 cm. The panicle first appears in the form of a bud. The time required for the appearance of such bud from seeding varies with different cultivars. For the bud to develop into a panicle it takes 8–15 d. The flowers are usually unisexual and monoecious. The male (staminate) and female (pistillate) flowers are borne on the same inflorescence. The staminate flowers are at the base in 3–16 flowered cymes and the pistillate ones, on the top in 1–7 flowered cymes, occupying about 50–70 and 30–50%, respectively (Cobley 1956; Purseglove 1968; Pandey 1976; Sundararaj and Thulasidas 1976; Kulkarni and Ramanamurthy 1977). Sometimes reverse case may be found, i.e., pistillate flowers at the base and staminate ones at the top (Hooker 1890).

1.4.5.1 Monoecy Versus Dioecy

Monoecy in castor bean is classified into two categories:

1. *Normal monoecious*: Pistillate flowers on the upper part of panicle and staminate flowers on the lower part.
2. *Interspersed monoecious*: Pistillate and staminate flowers interspersed along the entire rachis.

The proportion of pistillate and staminate flowers among panicles varies widely both within and among genotypes. It is also influenced considerably by environment (Zimmerman and Smith 1966 *fide* Milani and Nobrega 2013).

In normal monoecious cultivars, the percentage of pistillate flowers along the rachis is usually the highest on the first panicle, with a decreasing percentage on subsequently developed panicles. With the decrease in pistillate flowers, there is a proportional increase in the number of staminate flowers (Zimmerman and Smith 1966 *fide* Milani and Nobrega 2013). This intraplant variation is generally associated with the seasons. Female tendency is the highest in spring and early summer; male tendency is the highest in mid- and late summer. Temperature is probably the main environmental component affecting sex. Moderate temperature promotes female flowers while high temperature promotes male flowers. However, age of plant and nutrition may also influence sex expression. Femaleness is the strongest in young plants with a high level of nutrition while maleness is the strongest in old plants with a low level of nutrition (Shifriss 1956 *fide* Milani and Nobrega 2013).

In addition to monoecy, a subtype of dioecy occurs in plants with only pistillate flowers along the entire rachis of all panicles (Zimmerman and Smith 1966 *fide* Milani and Nobrega 2013). The counterpart, plants with only staminate flowers, occurs in extreme climatic conditions, with high temperature or water deficit (Fig. 1.2).

1.4.6 Flower

1.4.6.1 Male Flower

Pedicellate, pedicel 0.5–1.5 cm long; bracteate, ebracteolate, actinomorphic, incomplete, apetalous.

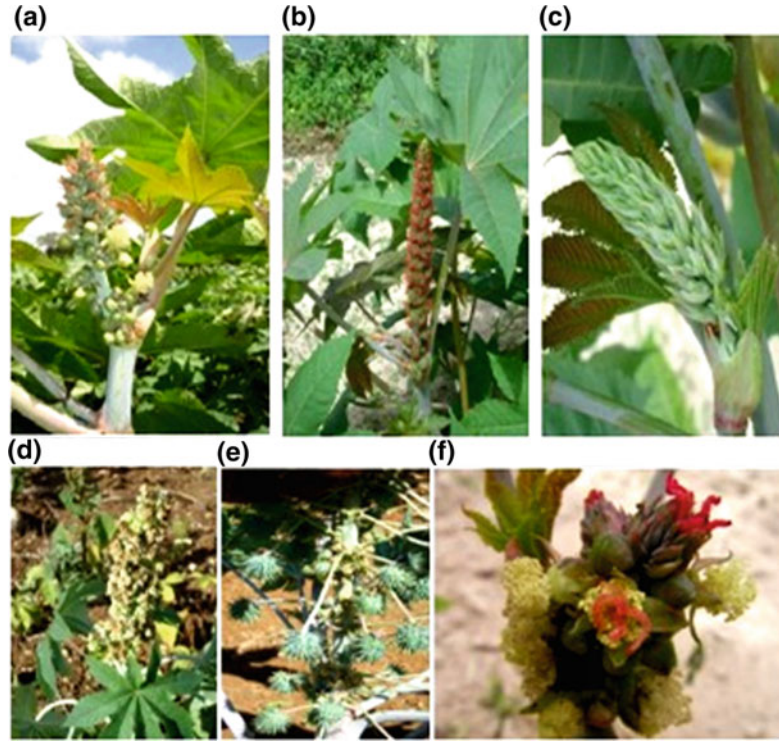

Fig. 1.2 Arrangements of male and female flowers in inflorescences of castor: **a** monoic normal; **b** and **c** gynodioic; **d** androdioic; **e** interspersed; **f** monoic bearing some perfect flowers. *Courtesy* Milani and Nobrega (2013)

Calyx	Sepals 3–5, ovate, spreading, 5–7 mm long, membranous, polysepalous, green, valvate estivation.
Corolla	Absent.
Androecium	Stamens 5; each filament profusely branched; anthers borne in ultimate branchlet, small, spherical, pale yellow, basifixed, introrse.

1.4.6.2 Female Flower

Pedicellate, pedicel 4–5 mm long; bracteate, ebracteolate, actinomorphic, incomplete, apetalous, hypogynous.

Calyx	Sepals 3–5, spathaceous, 3–5 mm long, membranous, connate and bursting irregularly, caducous (quick falling), green, valvate estivation.
Corolla	Absent.
Gynoecium	Ovary superior, covered with fleshy soft spines each terminating in a transparent bristle which breaks off as fruit develops, tricarpellary, syncarpous, trilocular, axile placentation, single large ovule in each locule; styles 3, very short; stigmas 3, deeply bifid, fleshy, red, papillate, long persistent.(Cobley1956; Purseglove 1968; Pandey 1976; Sundararaj and Thulasidas 1976, Shukla and Misra 1982)

1.4.7 Fruit

"Schizocarpic capsule" or regma with elongated pedicel, globose, 3-lobed, 1.5–2.5 cm in diameter, pericarp usually spiny, but may be smooth, warty or echinate; unripe fruits green, but red in

Fig. 1.3 Different colors in the fruits of castor. *Courtesy* Milani and Nobrega (2013)

Fig. 1.4 Developing fruits. *Courtesy* Wikipedia, the free encyclopedia

Fig. 1.5 Fruit dries and splits into three cocci, forcibly ejecting the seeds. *Courtesy* Wikipedia, the free encyclopedia

some cultivars, turning brown on ripening (Figs. 1.3, 1.4); woody remains of pistil at apex. Dehisce by woody pericarp splitting along dorsal suture into three cocci (Fig. 1.5); in wild plants and some cultivars splitting is violent ejecting seeds; most modern cultivars are indehiscent and seed is held in regma for several weeks (Cobley 1956; Purseglove 1968; Pandey 1976; Sundararaj and Thulasidas 1976; Shukla and Misra 1982; Nanda and Agrawal 2009).

Panicles may be conical, cylindrical, or oval in shape with different fruit arrangements. The

Fig. 1.6 Seeds of castor plant. *Courtesy* Wikipedia, the free encyclopedia

fruit arrangement may be compact, semicompact, or loose (Salihu et al. 2014).

1.4.8 Seed

Three seeds/regma; ovoid, elongated, oval or square in shape; compressed dorsally, tick-like, shining, pale gray, or pale buff to almost black with darker mottling (Fig. 1.6); carunculate, yellowish-white integumentary warty appendage called caruncle at micropylar end (base), a type of elaiosome, promoting dispersal of seed by ants (myrmecochory); hilum almost concealed under caruncle; size variable, 0.5–1.5 cm long; intercultivar and interpanicle variation in 100-seed weight 9–100 g; seed coat differentiated into testa and tegmen; testa (hull) brittle, about 20% weight of the seed; tegmen, thin integument enclosed by testa; raphe very prominent; albuminous (endospermic), endosperm copious, embryo small, cotyledons thin and papery; poisonous, allergenic (Cobley 1956; Purseglove 1968; Sundararaj and Thulasidas 1976; Kochhar 1981; Nanda and Agrawal 2009; Salihu et al. 2014).

1.4.9 Floral Formula

i. Male flower: Br Ebrl ⊕ ♂ $K_{3–5}$ C_0 A_5 (branched) G_0
ii. Female flower: Br Ebrl ⊕ ♀ $K_{3–5}$ C_0 A_0 $\underline{G}_{(3)}$

1.5 Floral Biology

The main shoot terminates in an inflorescence which is the oldest and those on the branches are later ones. It is generally believed that the male flowers open first and the female ones 1–2 d later. However, it has also been found that in many cases both male and female flowers open at the same time, and in a few cases either males or females open first (Sundararaj and Thulasidas 1976).

A male flower, after opening, sheds viable pollen grains for 1–2 d. The best weather conditions for pollen shedding are 26–29 °C temperature and 60% relative humidity (RH), which may vary according to cultivar. High temperature, plant age, and short-day length favor the appearance of male flowers (Milani and Nobrega 2013). Dehiscence of anthers takes place between 4 a.m. and 8 p.m. (Sundararaj and Thulasidas 1976). Pollen shedding occurs 2–3 h after sunrise until late afternoon with peak shedding at mid-morning. Pollen grains are viable for a long period ranging from 2 to 21 d with a viability of 76–26%, respectively, and up to one year under deep-freeze conditions or freeze drying (Hegde and Lavanya 2015). However, as per Sundararaj and Thulasidas (1976), the pollens remain viable for 66 h.

Regarding the female flowers, most researchers claim that the male flowers mature first and anthesis usually occurs in a short period of time before opening of the female flowers. The stigma becomes fully receptive in a few hours after opening of the flower. It remains receptive for a period of 5–10 d after anthesis depending on

environmental conditions. However, some researchers claim a short protogyny (Milani and Nobrega 2013). According to Hegde and Lavanya (2015), the stigma remains viable for 5–6 d and by tenth day, fertilization decreases to 9–10%. The exertion of the stigmatic lobes of the female flowers occurs at all hours, and it takes nearly 1–3 d for the stigmas to attain full development and become receptive from the time of opening of the female flowers. Stigmas remain receptive for 2–3 d (Sundararaj and Thulasidas 1976).

Pollen shedding (dehiscence of anthers) and viability, and receptivity of stigmas depend on cultivars and environmental conditions.

1.6 Pollination

The flowers of the central panicle tend to self-pollinate (30–70%), and those of lateral panicles are cross-pollinated (Hegde and Lavanya 2015). A fairly high percentage of self-pollination occurs due to the structure of the flowers and their disposition. But, cross-pollination also takes place (Sundararaj and Thulasidas 1976). So, castor bean has a mixed mating system generating both self- and cross-fertilized offsprings. Under natural conditions, cross-pollination in castor bean can exceed 80%, but the actual level of cross-pollination is dependent on both genotypes and environmental conditions (Milani and Nobrega 2013). A natural outcrossing of 5–14% occurs in castor bean. This is considered high in crop improvement work, and hence castor bean is a highly cross-pollinated crop (Sundararaj and Thulasidas 1976). Unlike other cross-pollinated crops, castor bean is inclined toward self-pollination with low inbreeding depression (Moshkin 1986).

It is usually stated that castor bean is mainly wind-pollinated. But as the flowers open, the glands on the young leaves on the sympodial branches below the inflorescence exude copious nectar. So, it seems probable that insects play part in pollination (Purseglove 1968). In large areas of the crop, pollen can be transported up to 2.5–3.0 km away from the originating plant (Hegde and Lavanya 2015).

1.7 Germination

Sound seed will retain their viability for 2–3 years. More even germination may be obtained by pouring boiling water on the seeds and leaving them to soak for 24 h. Some castor bean seeds show dormancy, but this can be broken by removing the caruncle and cutting a small hole on the testa. Germination is epigeal, and emergence takes 7–10 d; sometimes longer (Purseglove 1968).

Earlier it was believed that caruncle absorbed moisture and helped the seed germinate. Now, it is understood that this is not essential for the absorption of moisture. A cultivar from Persia is devoid of the caruncle but gives good germination (Sundararaj and Thulasidas 1976).

Suitable soil temperatures for germination are between 10 and 18 °C, and average day temperatures of 20–26 °C with a minimum of 15 °C and a maximum of 38 °C, with low humidity. The hard and brittle testa breaks near the caruncle and the radicle emerges. With a little more growth, the hypocotyl pulls out of the soil with the thin cotyledons still enclosed by the endosperm. The testa gets split to a great extent and falls off the mass of endosperm which is being absorbed by the growing cotyledons. The cotyledons make rapid growth absorbing nutrients from the endosperm, subsequently opening up into two green oblong or oval leafy structures. The remnants of the endosperm fall off as a thin mass. The growing plumule and the radicle get the prepared nutrients from this first pair of green seed leaves (Sundararaj and Thulasidas 1976; http://www.cabi.org/isc/datasheet/47618).

1.8 Anatomy

Very little work has been done on the anatomy of castor bean. Pentfound (1932) carried out anatomical studies on castor bean plants grown in different soils, moisture content and light-intensity combinations. The root system showed better development in sunlight than in the shadow, despite greater leaf area. The development of the mechanical system was

related to the intensity of light and soil moisture condition. Scott (1937) studied the development of spiral vessels in detail. The development and distribution of calcium oxalate in mature green plants, etiolated plants, green and etiolated seedlings was also studied by her (Scott 1941). Reynolds (1941) made a detailed study of the development of node and the differentiation of the vascular network in castor bean. Scott and Sharasmith (1933) studied the transition region in the seedling of castor bean and discussed the findings from a physiological standpoint. The following information is the outcome of the studies conducted at the Rajendranagar Farm (Kulkarni and Ramanamurthy 1977).

Root The root develops cork externally and the first phellogen originates in the cortex.

A single endodermal layer is present. The pericycle consists of strong thick-walled elements.

The xylem is tetrarch and is visible even after considerable secondary thickening. The secondary xylem is characterized by the presence of large-diameter vessels. The entire wood is roughly divided into four sectors by fibrous zones. The rays are uniseriate. The vascular elements appear embedded in a sheet of fibrous tissue when observed in a transverse (cross) section.

Stem The epidermis is single-layered with a cuticle, and is interrupted by large secretory cells which are filled in by oily or resinous substances. The epidermis shows an elevation over the secretory cells.

The hypodermis is collenchymatous in 8–9 layers and is interrupted by a cortical zone consisting of plates of chlorenchyma.

This collenchymatous hypodermis is also interrupted by the large secretory cells.

The chlorenchymatous cortex is five- or six-layered with intercellular spaces. The endoderm is not clear. The pericycle as such, consisting of fibrous cells, is not recognizable. The vascular system in the primary condition consists of conjoint, collateral bundles separated by medullary rays. The pith consists of polygonal parenchymatous cells.

John and Rao (1942) reported that the anatomy of the stem of seedlings of annual and perennial types revealed that the stem structure of perennial type differs from the annual one in that it has a more or less continuous cylinder of primary xylem ensheathed by an unbroken cylinder of secondary xylem which is added on indefinitely. In the annual ones, the vascular tissues are arranged in the form of discrete or discontinuous bundles in a cylinder.

Leaf The leaf is dorsiventral. The upper and lower epidermis are interrupted here and there by secretory cells which are flask-shaped. The palisade is single-layered with elongated and small closely abutting cells. The spongy tissue is relatively less interspaced.

In the transverse (cross) section of the petiole, there are as many as eight vascular bundles running all along the length and, in general, the structure is very similar to that of the stem.

1.9 Embryology

No mention is found in published literature about work on embryology of castor bean. Maheshwari (1950) made the following mention:

"Regarding the final fate of the endosperm in *Ricinus*, it forms a permanent storage tissue which persists until the germination of the seed, while in others it is used up by the growing embryo and is no longer seen in the mature seed."

1.10 Ecology

1.10.1 Climate

Castor bean is a hardy crop which survives in a wide range of environments. It is essentially a

warm season crop, cultivated in tropical, subtropical, and temperate regions. Its cultivation is largely confined to countries lying between 40°N and 40°S latitudes, but in Russia, a few varieties are grown even up to 52°N latitude. It is being cultivated up to an attitude of 2500 m with an optimum between 300 and 1500 m. But in regions where frosts are common during the crop season, its cultivation is restricted up to 500 m. A frost-free growing period of 140–190 d depending on variety is highly essential for obtaining satisfactory yield. It grows in tropical and subtropical regions as a perennial plant and in temperate climate as an annual. A moderate temperature of 20–26 °C is highly favorable during crop period for obtaining higher yield. A low temperature (less than 15 °C) in the seed bed prolongs the emergence of seedlings. Seed may fail to set if the temperature is above 38 °C for an extended period. Castor bean, being a deep rooted crop, is fairly resistant to drought. It is generally raised under rainfed conditions and a well-distributed rainfall of 500–600 mm during growing period is necessary for reasonably good yield. Castor bean can withstand long dry spells as well as heavy rains but is highly susceptible to water-logged conditions. Castor bean is basically a long-day plant, but is adaptable with less yields to a wide range of photoperiod. However, castor bean flowers normally on both a short 12 h and a long 18 h day, but at 9 h growth and development were severely retarded (http://www.cabi.org/isc/datasheet/47618; http://www.jatrophabiodiesel.org/castor/about-plant.php; Purseglove 1968; Kulkarni and Ramanamurthy 1977; Shiv Raj 1978; Nanda and Agrawal 2009; Salihu et al. 2014).

1.10.2 Soils

Castor bean grows almost anywhere land is available. It can be grown on a wide range of soils, provided they are fairly deep and well drained. Heavy clays, with poor drainage, and marshy soils are unsuitable. The highly suitable soils for castor bean are deep, moderately fertile, with slightly acidic conditions (pH 5.0–6.5), well-drained, sandy loams. Castor bean tolerates pH of 4.5–8.3. Castor bean can tolerate a pH of 8.0 but at this pH, the soil structure and physical properties become the important limiting factors for castor bean cultivation. It also tolerates dry-arid conditions and grows in semi-arid and arid soils with poor water-holding capacity. Excessively fertile soils are not desirable, as they favor excessive vegetative growth at the cost of seed yield. Insufficient nitrogen results in reduced seed yields. Excessive nitrogen results in extensive and heavy vegetative growth with non-significant increase in yield (http://www.cabi.org/isc/datasheet/47618; http://www.jatrophabiodiesel.org/castor/about-plant.php; Purseglove 1968; Kulkarni and Ramanamurthy 1977; Shiv Raj 1978; Nanda and Agrawal 2009; Salihu et al. 2014).

1.11 Chemical Composition

Seeds of commercial cultivars contain 40–55% of non-drying oil. Large-seeded cultivars have less oil than smaller-seeded ones. The oil contains 80–90% of ricinoleic acid, which is unique among vegetable oils in having a hydroxyl group near a double bond. On dehydration a drying oil is produced, which does not turn yellow on drying or baking; hence its value in paints and varnishes. The seeds contain a toxic protein, ricin, which acts as a blood coagulant. Eating a single seed may cause serious illness and more may be fatal. They contain a powerful allergen, a glycoprotein. Neither ricin nor the allergen is carried over into the oil if it is properly extracted, but they remain in the meal. The leaves contain ricin, but in much smaller quantities than in the seeds; they should not be fed to livestock (Purseglove 1968).

1.12 Economic Botany

1.12.1 Castor Oil in Medicine and Cosmetics

Castor oil is very effective laxative, and it is used as moisturizer for the treatment of skin problems

like sunburn, wrinkles and fine lines, dry skin and stretch marks. It has also been claimed that it prevents infections like warts, boils, athlete's foot, acne, ringworm, and chronic itching.

1.12.2 Castor Bean in Agriculture

Castor Bean Meal and Husk for Animal Feed:

Detoxified castor bean meal (castor cake, pomace, oilcake), the residue left after extraction of oil, can be used as animal feed. Castor bean meal detoxified by boiling could be added up to 100 g/kg in broiler finishing diets without any harmful effects. Castor bean meal detoxified by autoclaving can replace up to 67% of the soybean meal in sheep rations. The husk is a low value by-product that can be used as roughage for ruminants. A sample castor bean husks containing a considerable amount of seed fragment (60 g/kg) was evaluated for feeding dairy goat. When hay was completely replaced by castor bean husks, there was reduction (27%) in milk but increase (28%) in lipid concentration. The husks were not subjected to any detoxification process, and no symptom of toxicity was observed (Kochhar 1981, Salihu et al 2014).

Castor Bean Meal as an Organic Fertilizer:

The use of castor bean meal as organic fertilizer is very advantageous because of high N content, fast mineralization, and anti-nematode property. The mineralization of castor bean meal was found to be 7 times faster than bovine manure and 15 times faster than bagasse of sugarcane. Castor bean meal promotes the growth in wheat and castor bean plants. Castor bean husks can also be used as organic fertilizer but must be blended with an N-rich organic material to provide a better balanced nutrient for plant growth (Salihu et al 2014).

1.12.3 Castor Bean in Ornamental Horticulture

Castor bean is used extensively as a decorative plant in parks and other public places, particularly as a "dot plant" in traditional bedding schemes. If sown early, under glass, and kept at a temperature of around 20 °C until planted out, the castor bean plant can reach a height of 2–3 m in a year. In areas prone to frost, it is usually shorter, and grown as if it were an annual (Phillips and Rix 1999).

1.12.4 Castor Oil in Biodiesel

Production of biodiesel from castor oil is technically feasible. The major constraint has been the high price paid for the oil as industrial oil because of high demand by the chemical industries to manufacture very high-value products. Biodiesel produced from castor oil has a remarkable advantage regarding lubricity because of its high energy-value and positive fuel properties (Salihu et al. 2014).

1.12.5 Other Industrial Uses of Castor Bean

Castor oil can be used as bio-based polyol in the polyurethane industry. In food industry, castor oil is used in food additives, flavoring, candy, and as mold inhibitor. The oil can also be used to prevent rice, wheat, and pulses from rotting. The oil is also important raw material in paints and nylon industries. Castor bean wax produced by hydrogenation is used in polish, electrical condensers, carbon paper, and as a solid lubricant (Salihu et al. 2014).

1.12.6 Castor Bean Seed as Food Condiment

The white, large seeds of castor bean are important sources of food condiment called Ogiri in the South-Eastern part of Nigeria. The condiment is believed to improve eye sight. The condiment is prepared by: remove the seed coat —boil the cotyledons for 8–10 h—sieve the cotyledons—leave the sieved cotyledons to stand for 12–14 h—grind the cotyledons to paste as Ogiri. The condiment can be stored for several

months as the oil content inhibits the growth of microbes (Salihu et al. 2014).

References

Brar KS, Sandhu GS, Labana KS (1977) Note on the resistance of some castor varieties to *Empoasca flavescens* and *Dichocrocis punctiferalis*. Indian J Agric Res 11(1):48–50

Cobley LS (1956) An Introduction to the botany of tropical crops. Orient Longmans Pvt Ltd, Calcutta, India, pp 113–116 Reprint 1957

De Candolle (1860) Prodromous systematis naturalis regni vegetabilis. 17:1016

FAOSTAT (2014) http://www.fao.org/faostat/en/. Accessed on 15 Aug 2017)

Hegde DM, Lavanya C (2015) Castor (*Ricinus communis* L.). In: Bharadwaj DN (ed) Breeding of field crops. Agrobios (India), Jodhpur, India, pp 471–512

Hooker JD (1890) Flora of British India 5. L. Reeve and Co, London

http://waynesword.palomar.edu/plmar99.htm. Accessed on 21 June 2017

http://www.cabi.org/isc/datasheet/47618. Accessed on 6 July 2017

http://www.jatrophabiodiesel.org/castor/about-plant.php. Accessed on 6 July 2017

Hutchinson J (1959) Families of flowering plants, vol I, II, 2nd edn, Oxford, London

John CM, Rao UN (1942) Perennial or tree castor. Madras Agric J 30(4):109

Kochhar SL (1981) Economic botany in the tropics. MacMillan India Ltd, Delhi, India, pp 170–171

Kulkarni LG, Ramanamurthy GV (1977) Castor (Revised edition). Indian Council of Agricultural Research, New Delhi, India

Maheshwari P (1950) An introduction to the embryology of angiosperms. Tata McGraw Hill Publishing Co, New Delhi, p 257 Reprint 1982

Milani M, Nobrega MBM (2013) Castor breeding. In: Andersen SB (ed) Plant breeding from laboratories to fields. InTechOpen. http://www.intechopen.com/books/plant-breeding-fromlaboratories-to-fields, pp 239–254

Moshkin VA (1986) Castor. Amerind Publishing Co, New Delhi

Nanda JS, Agrawal PK (2009) Botany of field crops, vol I. Kalyani Publishers, Ludhiana-New Delhi, India, pp 303–307

Narain A (1952) Bloom character in castor oil plant, *Ricinus communis* L. Curr Sci 21(6):166

Pandey BP (1976) Taxonomy of angiosperms, 3rd edn. K Nath & Co, Meerut, India, pp 27–37, 219–225

Pentfound WT (1932) The anatomy of castor bean as conditioned by light intensity and soil moisture. Am J Bot 19:538–546

Phillips R, Rix M (1999) Annuals and biennials. MacMillan, London, p 106

Purseglove, JW (1968) Tropical crops: dicotyledons. vol 1 & 2 combined, 1st ELBS edn 1974. ELBS and Longman, UK, pp 180–186 Reprint 1984

Reynolds ME (1941) Development of the vascular network in node of the *Ricinus communis*. Am J Bot 28:728

Salihu BZ, Gana AK, Apuyor BO (2014) Castor oil plant (*Ricinus communis* L.): botany, ecology and uses. Int J Sci Res 3(5):1333–1341

Scott FM (1937) Differentiation of the spiral vessels in *Ricinus communis*. Bot Gaz 99:69

Scott FM (1941) Distribution of calcium oxalate crystals in relation to tissue differentiation and presence of other ergastic substances. Bot Gaz 103:225–246

Scott FM, Sharasmith HM (1933) The transition region in the seedlings of *Ricinus communis*—a physiological interpretation. Am J Bot 20:176

Seshadri CR, Seshu KA (1956) Observations on jassid injury in castor. Madras Agric J 43:197–199

Seshadri CR, Varisai M (1951) Study of inheritance of certain characters in castor. In: Proceedings of first scientific workers conference, Agricultural College & Research Institute, Coimbatore, Tamil Nadu, India

Shifriss O (1956) Sex instability in *Ricinus*. Genetics 41:265–280

Shiv Raj A (1978) An introduction to physiology of field crops. Oxford and IBH Publishing Co, New Delhi, pp 123–127 Reprint 1985

Shukla P, Misra SP (1982) An introduction to taxonomy of angiosperms, 3rd edn. Vikas Publishing House Pvt Ltd, New Delhi, India, pp 328–338

Sundararaj DD, Thulasidas G (1976) Botany of field crops. The MacMillan Co of India Ltd, Delhi, India, pp 264–273

Watt G (1892) A dictionary of economic products of India. 6(1):506–507

Weibel RO (1948) The castor-oil plant in the United States. Econ Bot 2(3):273–283

Weiss EA (1971) Castor, sesame and safflower. Barnes & Noble, New York

Weiss EA (2000) Oilseed crops, 2nd edn. Blackwell Science, Oxford

White OE (1918) Inheritance studies on castor beans. Mem Brooklyn Bot Gdn 1:513–521

Zimmerman LH, Smith JD (1966) Production of F_1 seed in castor bean by use of sex genes sensitive to environment. Crop Sci 6:406–409

Genetic Diversity in Castor Bean

2

Prathap Reddy Kallamadi and Mulpuri Sujatha

Abstract

Castor bean is an important nonedible oilseed crop that has many industrial uses. Genetic diversity in a germplasm is the fundamental requirement for crop improvement programs. There are several genetic markers available for assessment of genetic diversity among the genotypes and accessions. Though castor bean is a monotypic, it exhibits wide phenotypic diversity. In castor bean, genetic markers such as agro-morphological characters, biochemical and cytological markers were widely used in characterization of genetic variation in the germplasm from India, Nigeria, Turkey, China, Brazil, Iran, and Ethiopia which indicated a low-to-high-level diversity in the castor bean germplasm depending on the markers and the genotypes studied. Genetic diversity in castor bean was assessed by using both dominant and codominant molecular markers (random amplified polymorphic DNA, RAPD), inter-simple sequence repeats (ISSR), start codon targeted (SCoT), amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR), expressed sequence tag-simple sequence repeats (EST-SSR), and random microsatellite amplified polymorphic DNA (RMAPD), and also advanced molecular markers, such as single nucleotide polymorphism (SNP), sequence-related amplification polymorphism (SRAP), target region amplification polymorphism (TRAP), and methylation-sensitive amplification polymorphism (MSAP). Molecular marker utilization in the characterization of genetic diversity in germplasm from countries including India, China, Brazil, Mexico, and worldwide collection revealed a low-modest genetic diversity in castor bean. The availability of castor bean genome has accelerated the molecular marker development of easy, fast, and reliable DNA markers for the assessment of genetic variation in the castor bean and further these would serve as a major tool in the characterization of genetic diversity and assist marker-assisted breeding programs for crop improvement.`

2.1 Introduction

Germplasm collections constitute one of the world's most readily available sources of plant genetic material (Allard et al. 1991). Managers of these collections strive to accumulate and maintain these collections as a way of preserving the

P. R. Kallamadi (✉) · M. Sujatha
Crop Improvement Section, ICAR-Indian Institute of Oilseeds Research, Rajendranagar, Hyderabad 500 030, India
e-mail: prathapreddykallamadi@gmail.com

C. Kole and P. Rabinowicz (eds.), *The Castor Bean Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97280-0_2

biological diversity of crops and other economically important plant species (Gilbert et al. 1999). Genetic diversity is defined as the variation in genes and genotypes (Rao and Hodgkin 2002). An understanding of the extent of genetic variation among the germplasm collection is the prime prerequisite for success of a breeding program (Allard et al. 1991). The improvement and development of new varieties and hybrids purely depend on the extent of exploitable genetic variation available in the parental material. In castor bean, very wide phenotypic diversity is observed, which is due to its cross-pollinating nature and outcrossing (Allan et al. 2008). Before utilizing the germplasm resource, it is essential to characterize the germplasm with regard to the level of diversity. Genetic markers are available to assess the extent of genetic diversity in a germplasm and are mainly categorized into three major types, viz, morphological, biochemical, and molecular markers. The ideal genetic marker should possess the characteristics of being codominant for discrimination of homozygotes from heterozygotes, single copy, economical to use, highly polymorphic, easily assayed, multi-functional, robust, and genome-specific in nature. This chapter presents an account of the different type markers including morphological, biochemical, cytogenetical, and molecular used for the assessment of genetic diversity in castor bean.

2.2 Assessment of Genetic Diversity Using Morphological Markers

Plants exhibit variation in different morphological characters, which can be easily scored through visual observation. The morphological markers are generally used as they are rapid and cost-effective. However, the major limitations of these markers are that they are few in number and are subject to variation due to the influence of environment and developmental stages of plant. Castor bean exhibits enormous phenotypic plasticity, and hence several agro-morphological characters were used as markers for characterization of genetic variation among the genotypes. Castor bean is predominantly cultivated in India, Mozambique, China, Brazil, Myanmar, and Ethiopia (FAOSTAT 2016); and germplasm from India, Nigeria, Turkey, Iran, and Ethiopia was characterized by using morphological characters. The germplasm available at the Castor bean Germplasm Management Unit, Hyderabad, India was characterized not only for the agro-morphological traits but also for agrnomically desirable traits and reaction to major biotic stresses. Anjani and Reddy (2002) characterized the genetic variation among 89 accessions collected from Northeast India by using eight morphological traits, and these accessions were categorized into six clusters based on D^2 analysis. Anjani (2010a) reported the genetic diversity among 23 early-maturing genotypes by using 14 morphological traits. The multi-year analysis of these genotypes revealed a high level of genetic diversity. A nonsignificant G × E interaction was observed for the traits including 100-seed weight, oil content, number of main stem nodes, days to 50% flowering, and days to maturity. Anjani (2010b) assessed the genetic diversity among 20 wilt-resistant castor bean genotypes. Nine important agro-morphological characters were considered in the estimation of genetic diversity in these genotypes. The large distance reported among the groups showed a significant genetic diversity within the 20 wilt-resistant accessions for their quantitative traits.

Udoh and Abu (2016) reported phenotypic characterization of ten castor bean accessions collected from savannah agro-ecological zones of Nigeria using agro-morphological characters, and the variation in morphological characters including fruit texture and stem color was significant among these accessions. Chavan et al. (2012) reported genetic divergence among 40 castor bean genotypes (India and a few exotic germplasm) based on 14 morphological characters. The results showed that the main contributors to genetic divergence in these genotypes were plant height (38.21%) and the effective length of primary spike (10%). Ramesh et al. (2012) assessed genetic variation among 64 Indian genotypes of castor bean for nine morphological characters. The percentage of oil

content was the major contributor (24.8%) for total genetic diversity, followed by seed yield/plant (20.14%) and 100-seed weight (13.69%). Arslan (2012) reported genetic variation among 17 castor bean populations naturally grown in the eastern Mediterranean regions of Turkey. The morphological characters including stem color, seed color, days to maturity, plant height, main spike length, 100-seed weight, and seed yield per plant showed significant variability among the populations. Goodarzi et al. (2012) characterized the genetic variation in 12 Iranian castor bean accessions based on 32 agro-morphological traits. The study reported low genetic diversity among the accessions and about 75% of the accessions were clustered into a single group.

A coreset of castor bean representing 165 genotypes with wide variation in 14 morphological characters was developed in India (Sarada and Anjani 2013) and further this coreset was evaluated for their resistance to wilt, root rot gray mold disease, leaf hooper, and contents of fatty acids and oil (Anjani et al. 2018). A set of 24 castor bean genotypes from India were evaluated for the extent of diversity for eight morphological traits (Onkarappa et al. 2014). Number of capsules (30.29%), effective spike length (23.91%), and oil content (20.80%) contributed the maximum to the total genetic diversity. Alemaw et al. (2014) assessed the genetic diversity among 105 castor bean accessions from Ethiopia based on 12 morphological characters. Out these morphological characters studied in this investigation, eight were influenced by the environment and showed a significant G × E interaction. Alemaw (2016) characterized 102 castor bean accessions from Ethiopia for their diversity by using 12 morphological traits. Higher variation among the accessions was reported and the majority of total variation was mainly contributed by capsules per plant and seeds per plant.

A study was undertaken on yield components and root traits to find out the genetic variation among 27 castor bean accessions (Ramesh et al. 2014). The maximum contribution toward the total genetic diversity in these accessions was from seed yield (31.19%), total dry matter (23.36%), and 100-seed weight (13.6%). Vinay et al. (2017) investigated the genetic diversity among 51 castor bean genotypes. The major contributors to total genetic diversity were plant height up to the primary spike (20.24%), 100-seed weight (15.43%), and effective spike length (14.43%). The inter-cluster distance among the genotypes ranged from 31.64 to 40.31. Rukhsar et al. (2018) reported on genetic variability among 15 castor bean genotypes by using yield and its component traits using 13 agro-morphological characters. The genotypic and phenotypic coefficients of variation were reported high for number of capsules on the primary raceme. A positive association between the number of nodes up to primary raceme and total length of primary raceme with seed yield per plant was observed. Yadav and Anjani (2017) characterized 392 castor bean accessions (Andaman and Nicobar Islands, India, Brazil, Hungary, Italy, Nigeria, and USA) for their oil content and fatty acid composition. The oil content among the germplasm ranged from 38 to 54%, and there was no significant variation in oil content between Indian accessions and exotic germplasm. The ricinoleic acid content in the germplasm ranged from 71.15 to 93.68% with an average of 85.68%, representing a significant variation among the ricinoleic acid content. Lower levels of linoleic acid and linolenic acid were reported.

2.3 Assessment of Genetic Diversity Using Biochemical Markers

Biochemical markers reveal variation among the individuals by exhibiting polymorphism in certain proteins, viz, isozymes, seed-storage proteins and secondary metabolites. These markers are fairly cheap, rapid, and codominant. The limited use of biochemical markers is due to their low polymorphism, labor-intensive nature and they are few in number. Similar to morphological markers, these markers are influenced by the environment and developmental stage of the plant. Arslan (2012) investigated the genetic variation among 17 castor bean genotypes that

are naturally grown in the Mediterranean regions of Turkey. The sodium dedocyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) of seed-storage proteins revealed high homogeneity in the major protein bands. However, there was a variation in the major protein bands. The similarity coefficient among the genotypes ranged from 0.364 to 0.955. Wang et al. (2013a) reported genetic diversity among 37 castor bean varieties by using peroxidase isoforms. The number of peroxidase isozyme bands varied from 3 to 7, and the varieties were grouped into five major categories at similarity coefficient of 0.79.

2.4 Assessment of Genetic Diversity Using Cytogenetic Markers

Cytological marker assays were performed with chromosome-specific probes such as, fluorescent in situ hybridization (FISH) and Genomic in situ Hybridization (GISH). Vasconcelos et al. (2010) investigated the chromosome markers for the characterization of ten castor bean genotypes. The techniques like chromomycin/4,6-diam inido-2-phenylindole (CMA/DAPI) banding, silver impregnation, FISH using rDNA 5S, and 45S probes were used in the analysis. The mean chromosome length ranged from 1.19 to 2.12 µm. Variation was observed among the genotypes with regard to the nondistentions of chromosome A satellite.

2.5 Assessment of Genetic Diversity Using Molecular Markers

Molecular markers (also known as DNA markers) are extensively used in the genetic characterization and improvement of crop species. A number of marker techniques were developed to assess genetic diversity in plants. The discovery of the Polymerase Chain Reaction (PCR) to amplify short segments of DNA gave rise to a second generation of faster and less expensive PCR-based markers. The emerging oligonucleotide-based technologies derived from the use of hybridization arrays and sequence-based techniques have accelerated the genome-wide diversity analysis. Molecular markers are hybridization-based such as restriction fragment length polymorphism (RFLP) and PCR-based; but in castor bean, the marker techniques employed were predominantly PCR-based. In castor bean, the genetic diversity analysis was carried out by using several marker techniques (Table 2.1). The availability of the draft genome of castor bean gave access to an enormous amount of sequence information, which resulted in the development of new insights into castor bean genomics. The following DNA-based markers were used for the assessment of genetic diversity in castor bean.

2.5.1 Random Amplified Polymorphic DNA (RAPD)

RAPD technique involves amplification of any DNA segment by using a short oligodeoxynucleotide primer of arbitrary nucleotide sequence (Kahl 2001). These markers are dominant, considerably abundant with low-cost running per data point. The main advantages of RAPD markers are; technically easy, does not require prior sequence information about the organism and results in robust polymorphism (Fig. 2.1a). RAPDs were the general choice of marker in many plants, where sequence databases were not available. The main disadvantages of this marker system lie in its non-reproducibility across laboratories and often across runs, besides being dominant. Gajera et al. (2010) determined the genetic variation among 22 genotypes from India by using 30 RAPD markers. A total of 200 RAPD primers were initially screened for polymorphism, of which only 15% were polymorphic. The low proportion of polymorphism in the RAPD analysis was attributed to the small number of genotypes used in the study. The similarity coefficient among the 22 genotypes ranged from 0.53 to 0.91 and the polymorphic information content (PIC) ranged from 0.63 to 0.90 with a mean value of 0.82. Li et al. (2012) reported low polymorphism and associated

Table 2.1 Genetic diversity studies in castor bean using molecular markers

Marker used	Germplasm	Results	Reference
RAPD			
30 RAPD and 5 ISSR markers	22 genotypes from India	Similarity coefficient ranged from 0.53 to 0.91 representing moderate genetic diversity	Gajera et al. (2010)
55 RAPD markers	37 Chinese varieties	Low genetic diversity	Li et al. (2012)
8 RAPD markers	40 genotypes from Sloavakia	The diversity index (DI) values ranged from 0.621 to 0.896 with a mean of 0.798	Vivodik et al. (2014)
27 RAPD markers	13 genotypes from India	The average PIC and diversity index (DI) values were 0.784 and 0.798, respectively	Lakhani et al. (2015)
145 RAPD, 42 ISSR and 10 SCoT markers	33 accessions obtained from seven different geographical regions of the world	RAPD markers exhibited a relatively higher level of polymorphism (54%) as compared to ISSR (38%) and SCoT markers (21%). This study reported modest genetic diversity.	Reddy et al. (2015)
ISSR			
10 ISSR markers	39 Northeastern China genotypes	Total genetic diversity was 65.5% within the population and 34.5% among the population	Wang et al. (2013b)
16 ISSR markers	12 accessions from six regions of Iran	The genetic diversity within population and total were 0.121 and 0.232, respectively.	Goodarzi et al. (2015)
21 ISSR markers	27 Brazilian germplasm accessions	The PIC values ranged from 0.137 to 0.499 with a modest diversity among the germplasm.	Vasconcelos et al. (2016)
AFLP			
3 AFLP combinations and 9 SSRs	A set of 41 accessions representing five continents (North America, South America, Europe, Asia, and Africa, representing 39 countries)	Lower level of genetic diversity reported. He = 0.126 for AFLPs and 0.188 for SSRs	Allan et al. (2008)
4 AFLP combinations and 7 SSRs	82 populations from Chiapas, Mexico	Estimated a high level of genetic diversity (71%) in the population	Quintero et al. (2013)
21 AFLP combinations	27 genotypes adapted to semiarid conditions of Brazil	Reported an average of 70.8% polymorphism	Vasconcelos et al. (2016)
SSR			
11 SSR markers	38 Brazilian genotypes	Developed 11 new microsatellite markers and validated	Bajay et al. (2011)

(continued)

Table 2.1 (continued)

Marker used	Germplasm	Results	Reference
28 SSR markers	72 accessions (68 accessions from Korea, 2 from Taiwan and one each from Israel and Russia)	Developed novel microsatellite markers and validated	Seo et al. (2011)
45 SSR markers	144 inbred lines derived from castor bean coreset (134 lines from India, 3 lines from USA and one each from Hungary, Nigeria, Australia, Brazil)	Modest gene diversity (0.382) observed in the germplasm collection and no genetic structure in the inbred panel	Senthilvel et al. (2017)
14 SSR markers	27 inbred lines from India	The Nei's genetic distance ranged from 0.04 to 0.62 with mean of 0.34	Rukhsar et al. (2017)
5 SSR markers	60 genotypes obtained from Slovakia	Average diversity index of 0.826.	Vivodik et al. (2016)
EST-SSR			
118 EST-SSR markers	24 castor bean germplasm accessions representing 14 countries	Genetic relationships between germplasm was moderate	Qiu et al. (2010)
135 EST-SSR markers	Nine Indian commercial castor bean hybrids and their parents	Development of EST-SSRs and assessment of genetic purity hybrids	Pranavi et al. (2011)
10 EST-SSR markers	14 genotypes from India	The estimated genetic variation is low	Ramana Rao et al. (2012)
35 EST-SSR markers	51 accessions from Northeast India	The mean PIC and discriminating power (DP) values were 0.38 and 0.59, respectively	Kanti et al. (2014)
29 EST-SSR markers	33 accessions collected from Andaman and Nicobar Islands	The average expected heterogeneity in the germplasm was 0.58, representing significant diversity among the collected germplasm.	Kanti et al. (2015)
28 EST-SSR primers	27 Indian accessions	Moderate genetic diversity	Ramesh et al. (2016)
22 EST-SSR markers	574 accessions of US germplasm collection (48 countries)	Moderate genetic diversity index of 0.53	Wang et al. (2017)
SNP			
48 SNP loci	152 accessions collected from USDA germplasm resources information network (represents worldwide distribution from 45 countries)	Genetic diversity low with mean observed heterozygosity across population of 0.15 and estimated heterozygosity of 0.21.	Foster et al. (2010)

(continued)

Table 2.1 (continued)

Marker used	Germplasm	Results	Reference
SRAP			
20 SRAP markers	81 Chinese varieties	The genetic similarities ranged from 0.32 to 0.92.	Lu et al. (2010)
SRAP markers	50 Chinese accessions	Similarity coefficients ranged from 0.64 to 0.97.	Mei-Lian et al. (2012)
168 TRAP markers	40 genotypes developed at NBIO-UFRB, Brazil	A significant level of genetic dissimilarity (0.66).	Simoes et al. (2017)
RMAPD			
9 AP-PCR and 84 RMAPD markers	31 accessions from domestic regions of China.	At a heritable coefficient of 0.56, the 31 germplasm accessions were categorized into three groups.	Dong et al. (2012)
MASP			
20 MASP markers	60 landraces collected from 29 countries	The DNA methylation level ranged from 3.8 to 34.3% and the polymorphic methylated loci occurred in nuclear and organelle genomes of the castor bean landraces	He et al. (2017)

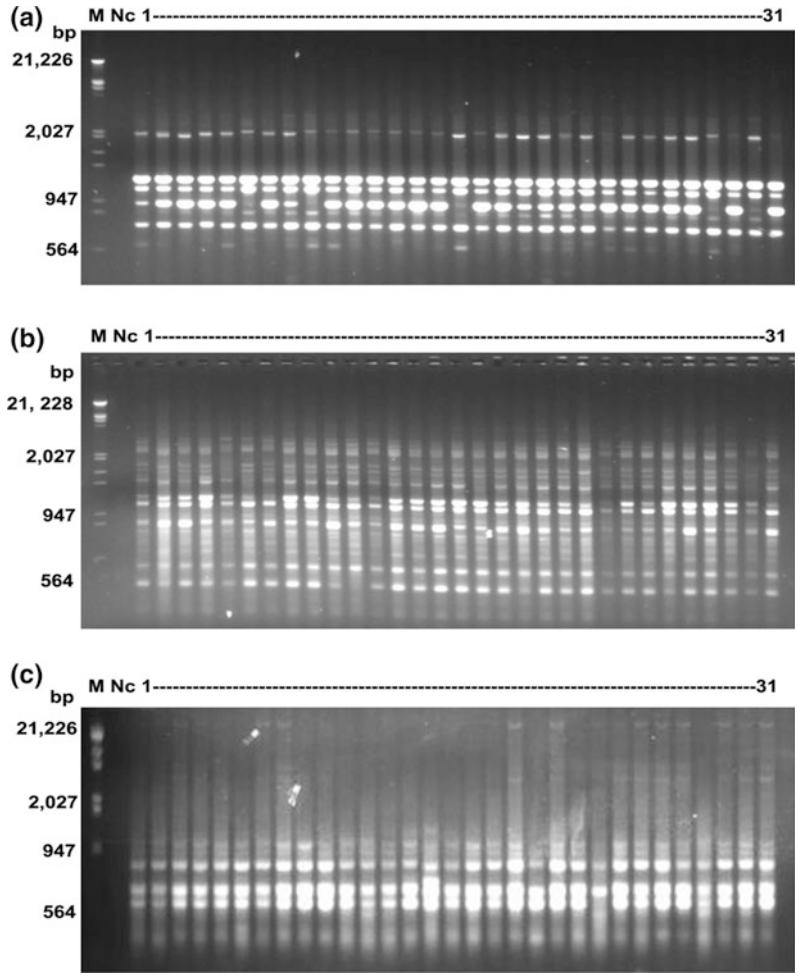

Fig. 2.1 **a** Electrophoretic analysis of DNA amplification produced using RAPD primer OPO-18. **b** Electrophoretic analysis of DNA amplification produced using ISSR primer UBC-856. **c** Electrophoretic analysis of DNA amplification produced using SCoT primer SCoT-14. Lanes designated as M represents DNA double digest with *Eco*RI and *Hind*III restriction enzymes, Nc-negative (no DNA) control, and 1–31 represent the DNA samples of 31 castor bean genotypes (Reddy et al. 2015)

genetic variation among the 37 Chinese castor bean varieties using 50 RAPD primers.

A study was undertaken to measure the genetic relatedness between 40 genotypes from Slovakia using eight RAPD markers (Vivodik et al. 2014). The diversity index (DI) value ranged from 0.621 to 0.896 with a mean of 0.798. Lakhani et al. (2015) performed RAPD analysis to determine genetic variation in 13 genotypes from India. A total of 27 RAPD primers were used in the assessment of genetic diversity, and the PIC of the primers ranged from 0.423 to 0.883 with an average of 0.705. Reddy et al. (2015) employed RAPD, inter-simple sequence repeat (ISSR) and start codon targeted (SCoT) markers for assessment of genetic variation in 31 castor bean accessions collected from seven geographical regions of the world. RAPD markers exhibited a relatively higher level of polymorphism (54%) as compared to ISSR markers (38%) and SCoT markers (21%). A total of 145 primers were found to be polymorphic out of the 214 primers screened. The high proportion of polymorphic primers (67%) was reported with PIC ranging from 0.06 to 0.49. The mean effective multiplexing ratio (EMR), marker index (MI), and resolving power (RP) were 2.33, 0.07, and 4.33 per primer, respectively. This study reported a modest genetic diversity with RAPD markers, and the similarity coefficient ranged from 0.47 to 0.91.

2.5.2 Inter-Simple Sequence Repeat (ISSR) Markers

ISSR markers are PCR-based markers, which use simple sequence repeat primers to amplify regions between their target sequences (Kahl 2001). This PCR-based technique exploits the abundant distribution of simple sequence repeats that are randomly present in the genome. The main advantages of ISSR technique are their low development costs, high portability across the crops, highly polymorphic and robust in usage (Fig. 2.1b). These markers are dominant that is the main disadvantage of this marker system. Five ISSR primers were used to investigate the molecular diversity among 21 castor bean genotypes from India (Gajera et al. 2010). This study reported a similarity coefficient range of, 0.51–0.93 among the genotypes. Wang et al. (2013b) employed ISSR marker technique to characterize the genetic diversity among 39 Northeastern China genotypes. The highest mean number of alleles for population belongs to Mongolian province, while the lowest for population from the Jilin province. The dendrogram analysis divided the 39 genotypes into four major clusters with a common node at 66% similarity. It is reported that 65.5% of the total genetic diversity were within the population and 34.5% was among the populations and these investigations revealed wide genetic diversity in the castor bean genotypes of Northeastern China.

Goodarzi et al. (2015) investigated the genetic diversity within the 12 castor bean accessions from six regions of Iran. The genetic diversity within population and total variation were 0.121 and 0.232, respectively. The AMOVA results showed that genetic diversity was mainly within an accession (82%) than among the accessions (18%). Reddy et al. (2015) assessed genetic diversity of 31 genotypes collected from seven geographical regions of the world using ISSR markers. The results showed the mean EMR, PIC, and RP were 1.34, 0.20 and 4.79 per primer, respectively. This study had reported a modest genetic diversity with ISSR markers and the similarity coefficient ranged from 0.34 to 0.94 and had reported few accessions specific bands, which can be converted into a SCAR marker for the identification of specific accessions. Vasconcelos et al. (2016) employed ISSR markers to find out the genetic diversity of Brazilian germplasm. A set of 21 ISSR primers had generated 264 loci, of which 187 were polymorphic. The PIC values in this study ranged from 0.137 to 0.499 indicating a modest diversity among the germplasm.

2.5.3 Start Codon Targeted (SCoT) Markers

Start codon targeted (SCoT) polymorphism is a PCR-based marker technique that involves amplification of the DNA region, surrounding the translation start codon (ATG). The technique employs a single primer as both forward and reverse primer similar to the RAPD and ISSR methods and is designed from the short-conserved regions flanking the start codon (ATG) of plant genes (Joshi et al. 1997; Sawant et al. 1999; Collard and Mckill 2009). These are functional markers that were widely used for genetic diversity studies and quantitative trait loci (QTL) mapping. Compared to the RAPD and ISSR techniques, SCoT polymorphism is reproducible but the markers are dominant like RAPD and ISSR markers (Fig. 2.1c). A study was reported on the use of SCoT markers for the characterization of 31 castor bean genotypes from different countries (Reddy et al. 2015). A set of 36 primers was employed in this study, of which ten primers generated polymorphism. The polymorphism of the primers ranged from 6 to 37% and the mean PIC, EMR, MI, and RP were 0.24, 0.61, 0.07, and 2.19, respectively. Similarity coefficient among the genotypes ranged from 0.41 to 1.0 and disclosed a modest level of genetic diversity through SCoT primers.

2.5.4 Amplified Fragment Length Polymorphism (AFLP) Markers

The AFLP technique combines the advantages of both restriction digestion and PCR amplification of the restriction-digested products (Vos et al. 1995). It is a stepwise procedure of digestion of genomic DNA with restriction enzymes (usually a frequent and a rare cutter) followed by ligation adaptors of a known sequence of the digested fragments. Then primers' complementary to the adaptors are used to amplify the restricted fragments by PCR. AFLP markers produce high polymorphism and are dominant in nature. The advantage of the technique is that it does not require prior sequence information and can be adapted with cDNA-AFLP. The disadvantage of AFLP is that it is technically challenging and a dominant marker system. Assessment of genetic variation in worldwide germplasm of castor bean using AFLP polymorphism is reported by Allan et al. (2008). This is the first report on utilization of molecular markers in assessment of genetic diversity of castor bean. A set of 41 castor bean accessions representing five continents (North America, South America, Europe, Asia, and Africa, a total of 39 countries) was subjected to molecular characterization. The study used three AFLP primer pairs, which generated a total of 303 loci (among those 43% were polymorphic). The minimum and maximum genetic diversity in between the accessions reported were 0.048 and 0.499, respectively, and the mean genetic diversity among the accessions was 0.294. The lower level of genetic diversity reported in the study might be due to the use of very few number of AFLP primers. AMOVA results showed that the majority of the genetic variation was detected within accessions (55%) than among the accessions (45%).

Quintero et al. (2013) employed AFLP technique for the assessment of genetic diversity among 82 populations from Chiapas, Mexico. The AFLP analysis of four combinations had reported an overall similarity coefficient that ranged from 0.03 to 1.0 and estimated a high level of genetic diversity (71%) in the population. Vasconcelos et al. (2016) reported an average of 70.8% polymorphism in 27 castor bean genotypes adapted to semiarid conditions of Brazil. A total of 278 loci were polymorphic (43.4%) fragments out of a total of 640 fragments generated by 21 AFLP primer combinations. AMOVA analysis reported higher variation within the groups (89.2%) than between the groups (10.8%).

2.5.5 Simple Sequence Repeat (SSR) Markers

Simple sequence repeats, also known as microsatellites, are the di-, tri-, tetranucleotide repetitive DNA sequences present in the genome of the organism (Tautz 1989). The SSRs are distributed abundantly in plants and flanked by conserved nucleotide sequences. These conserved nucleotide sequences are used from which the forward and reverse primers are synthesized to amplify the DNA regions containing SSR alleles by PCR. SSRs are codominant markers and have several advantages, i.e., rapid, highly polymorphic, multi-allelic and thus, making multiplexing possible. The main drawbacks of these markers are the high development and startup costs and normally give single loci even in polyploids. Allan et al. (2008) characterized the genetic variation among the worldwide germplasm accessions by using SSRs. A total of 41 castor bean accession representing five continents from 39 countries were subjected to characterization. The PIC of nine primer pairs used in the study ranged from 0.078 to 0.647. The minimum and maximum genetic diversity reported were 0.443 and 0.595, respectively, and the mean genetic diversity among the accessions was 0.595. The lower level of genetic diversity reported in the study was attributed to the limited

number of SSR primers. Bajay et al. (2011) developed 11 microsatellite markers to investigate genetic diversity in 38 Brazilian castor bean genotypes. The mean observed and expected heterogeneity were 0.060 and 0.416, respectively.

Seo et al. (2011) developed and used 28 polymorphic SSRs for the characterization of 72 castor bean accessions (68 accessions from Korea, two from Taiwan and one each from Israel and Russia). A total of 625 unigene clones were screened for microsatellites of which 423 clones (67.7%) contained SSRs. Of these identified clones, 340 (80.4%) were dinucleotide repeat motifs and the rest were trinucleotide repeat motifs. After screening 237 primers for polymorphism, only 28 were found to be polymorphic. The observed and expected heterogeneity mean values among the accessions were 0.01 and 0.31, respectively. The phylogenetic analysis showed three major clusters with a genetic distance of 0.23 suggesting low genetic diversity among the accessions. Quintero et al. (2013) reported the genetic relationships among 82 castor bean accessions from Chiapas, Mexico by using SSR and AFLP markers. Seven SSRs were used and the PIC values of the reported markers ranged from 0.5 to 0.8 and indices of genetic diversity ranged from 66 to 94% with an average of 70%. The presence of a mixture of genotypes from different regions in the cluster analysis indicated that there was a narrow geographic structure in the accessions.

A set of 144 inbred lines derived from castor bean coreset (134 lines from India, three lines from the USA and one each from Hungary, Nigeria, Australia, Brazil) was used to determine the genetic diversity and population structure by using 45 SSRs (Senthilvel et al. 2017). Lower allelic diversity reported in this study was due to use of 91% of SSR primers that belonged to trinucleotide repeats. The mean PIC and heterogeneity for the markers were 0.36 and 0.04, respectively. Modest gene diversity (0.382) was observed in the germplasm collection. The genetic structure of the population estimated by neighbor joining clusters, PCoA and STRUCTURE reported the lack of distinct genetic structure in the inbred panel. Rukhsar et al. (2017) reported a combined use of 13 morphological descriptors and 14 SSR markers in the assessment of 27 castor bean inbred lines from India. The PIC values ranged from 0.16 to 0.68 with an average of 0.43. The Nei's genetic distance among the 27 genotypes ranged from 0.04 to 0.62 with a mean of 0.34. Vivodik et al. (2016) investigated the genetic diversity among 60 genotypes obtained from Slovakia using five SSR primers. This study reported an average diversity index of 0.82.

2.5.6 Expressed Sequence Tag-Simple Sequence Repeat (EST-SSR) Markers

EST-SSRs are the simple sequence repeats that are derived from transcribed regions of the genome. Compared to genomic SSRs, EST-SSRs are three times more transferable across the species. The main limitations of EST-SSR markers are that they require sequence information and they also exhibit substantially decreased level of polymorphism. The advantages of these markers lie in their rapidity, highly reproducibility, and reliability. Qiu et al. (2010) characterized the genetic variation among 24 castor bean accessions representing 14 countries using EST-SSR markers. The analysis of genetic relationships among the germplasm accessions was moderate. The study revealed 118 polymorphic EST-SSR markers with 2.97 alleles per marker, and these markers displayed PIC and gene diversity values of 0.36 and 0.41, respectively. Pranavi et al. (2011) developed a set of 130 EST-SSR markers for assessment of genetic purity of nine Indian commercial castor bean hybrids. Ramana Rao et al. (2012) reported the results of genetic variation among the drought-tolerant and -sensitive genotypes by using 10 EST-SSR markers. The estimated genetic variation was low.

A total of 51 castor bean accessions from Northeast India were characterized for their genetic variation by using 35 EST-SSRs (Kanti et al. 2014). The mean PIC and discriminating power (DP) values were 0.38 and 0.59,

respectively. Kanti et al. (2015) assessed genetic variation among 33 castor bean germplasm lines collected from Andaman and Nicobar Islands, India by using agro-morphological traits and EST-SSRs. The total germplasm fell into three major groups. The average expected heterogeneity in the germplasm was 0.58, which indicated a significant diversity among the collected germplasm. Ramesh et al. (2016) investigated genetic variation among 27 Indian castor bean accessions by utilizing 28 EST-SSR primer pairs of which 12 primers showed polymorphism with an average PIC of 0.33. Wang et al. (2017) determined the genetic variation and population structure in 574 accessions of US germplasm collection with 22 polymorphic EST-SSR markers. The mean PIC (0.44) reported was slightly higher than the reported values of Senthilvel et al. (2017) and Allan et al. (2008), which were 0.403 and 0.36, respectively. This study reported a moderate genetic diversity index of 0.53. This is the only study which used a large number of accessions from 48 countries in determining the genetic diversity. The population was broadly divided into four subpopulations based on structure analysis. Genetic diversity among the subgroups were G_1 (0.49), G_2 (0.54), G_3 (0.39), and G_4 (0.39). The genetic distance (Nei's minimum distance) between the groups were reported as 0.22 (G_1 and G_2), 0.23 (G_1 and G_3), 0.17 (G_1 and G_4), 0.25 (G_2 and G_4), and 0.12 (G_3 and G_4).

2.5.7 Single Nucleotide Polymorphism (SNP) Markers

A single nucleotide polymorphism (SNP) refers to a variation at a single position in DNA sequences between individuals. SNPs are the most abundant polymorphic markers with a frequency of 2–3 polymorphic sites per kilobase of the genome (Cooper et al. 1985). The SNP data is highly dependent on sequence information. Several sequencing and bioinformatics methods are available for the identification and development of SNPs. These markers are robust in usage and highly suitable for high-throughput genotyping. However, the high development cost is the major limitation of this technique. Foster et al. (2010) used high-throughput SNP genotyping to assess the genome-wide variation and population structure in 152 castor bean accessions collected from USDA germplasm resources information network (total samples represent worldwide distribution from 45 countries). This study reported a low level of genetic diversity in the genotypes studied. The mean observed heterozygosity across the population was 0.15 and estimated heterozygosity was 0.21. The AMOVA results showed the majority of molecular variation within populations (74%) followed by among the population (22%) and among the continents (4%).

2.5.8 Sequence-Related Amplification Polymorphism (SRAP) Markers

Li and Quiros (2001) developed the marker technique known as sequence-related amplification polymorphism (SRAP). The objective of this marker method is for the amplification of open-reading frames (ORFs). It is a two-primer marker technique that consists of forward and reverse primers. The forward primer is of 17 base length, of which the first 5′ ten bases consist of a core sequence (filler sequence) followed by the sequence CCGG and then by three selective nucleotides at 3′ end. These markers have high reproducibility with the optimal marker distribution. The reverse primer is structurally similar to the forward primer except for AATT instead of the CCGG sequence. A set of 20 SRAP primers was used in the assessment of genetic relationships among 81 castor bean Chinese varieties (Lu et al. 2010). A total of 263 fragments were generated by 20 SRAP primers, of which 214 (81.37%) were polymorphic. The genetic similarities of 81 varieties ranged from 0.32 to 0.92. The cluster analysis grouped the varieties based on geographical regions. Mei-Lian et al. (2012) assessed the genetic diversity in 50 Chinese

accessions using the SRAP markers. The results indicated a lower level of polymorphism (29.97%) with the average gene diversity index of 0.09. Similarity coefficients among the germplasm ranged from 0.64 to 0.97.

2.5.9 Target Region Amplification Polymorphism (TRAP) Markers

TRAP is a technique based on the combination of conserved sequences derived from ESTs and arbitrary primers (Hu and Vick 2003). The TRAP marker method adapted the concept of SRAP in designing arbitrary primers with slight modification in the principles of primer design. The principles involved in designing the primers include: (1) The selective nucleotide, 3–4 nucleotides at the end, (2) The core, 4–6 nucleotides with AT- or GC-rich regions, (3) The filler sequences that make the 5′ end. The advantages of this method are that the assays are rapid with high reproducibility. Simoes et al. (2017) employed TRAP marker technique to estimate the genetic diversity in 40 castor bean genotypes developed at NBIO-UFRB, Brazil. A total of 168 primers were selected for amplification by using 28 fixed and six arbitrary primer combinations. The fixed primers were designed to target the regions of either the metabolic pathway of ricin or the fatty acid synthesis. A total of 64 effective TRAP primer combinations generated a total of 580 fragments, of which 335 (58%) fragments were polymorphic. The PIC of the TRAP markers ranged from 0.03 to 0.33 with a mean value of 0.24. A significant level of genetic dissimilarity (0.66) was reported among the castor bean elite lineage evaluated.

2.5.10 Random Microsatellite Amplified Polymorphic DNA (RMAPD) Markers

RMAPD is a PCR-based method which involves the use of a combination of random primers (RAPD) and microsatellite primers as forward and reverse primers (Lan et al. 2006). It has wide applications in determination of genetic diversity, population structure, and marker-assisted selection. There is a report on utilization of the AP-PCR (or RAPD) and RMAPD technique for the characterization of genetic diversity among 31 castor bean germplasm accessions from domestic regions of China (Dong et al. 2012). In this study, a set of 84 RMAPD and nine AP-PCR primer pairs were used for analyzing the genetic diversity. The number of polymorphic bands in AP-PCR ranged from 2 to 10, whereas in RMAPD, it ranged from 5 to 12. At a heritable coefficient of 0.56, the 31 germplasm accessions were categorized into three groups.

2.5.11 Methylation-Sensitive Amplification Polymorphism (MASP) Markers

DNA methylation is an important epigenetic mechanism that controls the gene expression, transposon mobility and genome integrity (Law et al. 2010). In plants, DNA methylation mechanism usually occurs in three sequence contexts; CG, CHG, and CHH (H might be A, C, or T). MSAP is a method used to detect the DNA methylation status of an organism. The differences in the methylation pattern between two individuals produces a differential polymorphism when digested with restriction enzymes, which target this region (Law et al. 2010). The MSAP technique involves a sequence of steps; (1) digestion of DNA with a methylation-sensitive restriction enzyme; (2) DNA fragments are ligated with adaptors in order to facilitate their amplification; (3) these generated fragments are selectively amplified by fluorescently labeled primers. This technique is highly reproducible and reliable.

MASP technique was employed to survey the genetic diversity in 60 landraces of castor bean collected from 29 countries (USDA National Plant Germplasm System and landraces from China). A reproducible PCR fragments are

achieved with 20 primer combinations sets and amplified fragments among the landraces ranged from 123 to 292. The DNA methylation level ranged from 3.8 to 34.3% in the 60 castor bean landraces. The study clearly demonstrated that the rate of DNA methylation greatly varies among different landraces (He et al. 2017). The percentage of polymorphic loci (PPB), number of different alleles (Na), number of effective alleles (Ne), Shannons information Index (I), and unbiased expected Heterozygosity (He) were observed as 69.07%, 1.69, 0.36, 1.39, and 0.24, respectively. The polymorphic methylated loci in the study were detected predominantly in the nuclear and organelle genomes of the castor bean landraces.

2.6 Transferability of Castor Bean Molecular Markers Across Taxa and Genera

The conservation of the structure of genetic regions across the genomes is well documented. The availability of genome sequences in the public domain and bioinformatics tools provide an alternative method for the identification of molecular markers (particularly SSRs). These identified molecular markers can be tested for their amplification across taxa, genera, and species. It reduces the developmental costs and saves time. Sharma and Chauhan (2011) showed higher rates of transferability (70%) of castor-bean-derived SSRs to 49 *Jatropha curcas* genotypes and eight *Jatropha* species. The genome analysis identified 5,80,986 SSRs in castor bean of which 302 randomly selected SSRs were tested for their cross transferability with *J. curcas* and eight other *Jatropha* species. Out of 302 castor bean SSRs tested, 211 primers showed amplification with *Jatropha* species. The transferability across the *Jatropha* species ranged from 58% (*J. gandulifera*) to 70% (*J. curcas*) with transferability of 63–68% with other species. A total of 37 primer pairs generated amplification in *J. curcas* and other species. Most of the polymorphic primers (35) belonged to dinucleotide repeats. Tetra- and pentanucleotide repeats failed to show polymorphism.

Qiu et al. (2010) tested the cross transferability of castor bean EST-SSRs to *Spernskia cantonensis* and *Jatropha curcas.* A total of 308 castor-bean-derived EST-SSRs were tested for their cross transferability with *S. cantonensis* and *J. curcas* of which 155 (50.2%) primer pairs amplified in *S. cantonesis* and 74 (24%) in *J. curcas.* Raji et al. (2009) tested cross transferability of EST-SSRs developed in cassava with castor bean. The results showed lower transferability of cassava derived EST-SSRs to castor bean. It is reported that only 13 (15%) out of 85 EST-SSRs were amplified in castor bean.

2.7 Genomic Resources Available in Castor Bean for Marker Development and Diversity Analysis

Arabidopsis is the first plant genome that was sequenced in the year 2000. After ten years of its publication, castor bean genome was published (First in the family of Euphorbiaceae) in the year 2010 by Chan et al. (2010). Before the publication of castor bean genome, there were only two reports on molecular marker utilization in castor bean (Allan et al. 2008; Foster et al. 2010). Availability of the castor bean genome in the public domain gathered a momentum in the development of molecular markers (particularly SSRs) and its use in the diversity assessment of the crop (Qiu et al. 2010; Bajay et al. 2011; Seo et al. 2011; Ramana Rao et al. 2012; Quintero et al. 2013; Ramesh et al. 2016; Wang et al. 2017; Senthilvel et al. 2017; Rukhsar et al. 2017; Vivodik et al. 2016). The availability of the genome sequence of castor bean and its genomic resources (Table 2.2) had accelerated the structural and functional genomics approaches to

Table 2.2 Genomic resources in castor bean

Database	Number
Genome assembly information	1
Genome survey sequences (GSS)	59
Nucleotide (DNA and RNA sequences)	65,942
High-throughput DNA and RNA sequence read archives (SRA)	1272
Expressed sequence tags	65,942
Collected information about gene loci	23,170

Source NCBI, accessed on 3rd April, 2018

discover the molecular genetic diversity and mapping important traits in the crop (Rivarola et al. 2011; Liu et al. 2016; Tan et al. 2016; Tomar et al. 2017).

2.8 Conclusion and Future Prospects

Genetic diversity studies in castor bean germplasm using morphological markers showed a high genetic variation in Indian and Ethopia and low-to-high in Nigeria, Turkey, China, Brazil, and Iran. The morphological characters such as 100-seed weight, oil content, number of main stem nodes, days to 50% flowering, days to maturity, the effective length of spike, and plant height were major contributors of morphological diversity existed across the germplasm from different countries. Molecular diversity indicated low-moderate genetic diversity in germplasm collection from India, China, Brazil, Mexico, and worldwide collection. The studies, which employed worldwide germplasm collection and used different marker systems had showed a narrow genetic diversity among the world collection and indicated a modest genetic variation among the Indian coreset germplasm. Morphological diversity in most cases was not in agreement with molecular diversity indicating a strong influence of environment in the assessment of genetic diversity of morphological characters. In few of the molecular diversity studies reporting high genetic diversity among the castor bean germplasm could either be due to use of low number of markers or limited number of genotypes. There is a need to combine both morphological and molecular markers to identify the variation in the germplasm and pooling up large germplasm sources in one place. These combined enhancements (morphological and molecular) would allow accurate assessment of the genetic diversity with a single collection representing divergent germplasm from worldwide. The availability of the castor bean genome had opened up a new avenue for development of novel molecular markers and also to construct dense-linkage maps for castor bean. The availability of genetic diversity information and dense-linkage maps in castor bean would provide the identification of molecular markers that are closely associated with important agronomical traits.

References

Alemaw G (2016) Diversity of castor (*Ricinus communis* L.) in Ethiopia. Ethiop J Agric Sci 26:57–67

Alemaw G, Kassabun BM, Taye G, Endalamaco C (2014) Phenotypic variability in Ethiopian castor (*Ricinus commnis* L.) accession. Int J Adv Biol Biomed Res 2:2909–2914

Allan G, Willams A, Pablo DR, Chan AP, Ravel J, Kein P (2008) Worldwide genotyping of castor bean germplasm (*Ricinus communis* L.) using AFLP and SSRs. Genet Resour Crop Evol 55:365–378

Allard RM, Alvim PDT, Ashri A, Barton JH (1991) Managing global genetic resources, the U.S. national plant germplasm system, elements of the national plant germplasm system. The National Academies Press, Washington DC, pp 43–86

Anjani K (2010a) Extra-early maturing germplasm for utilization in castor improvement. Ind Crops Prod 31 (1):139–144

Anjani K (2010b) Pattern of genetic diversity among *Fusarium* wilt resistant castor germplasm accession (*Ricinus communis* L.). Electron J Plant Breed 1:182–187

Anjani K, Raoof MA, Prasad MSL, Duraimurugan P, Lucose C, Yadav P, Prasad RD, Lal JJ, Sarada C (2018) Trait-specific accessions in global castor (*Ricinus communis* L.) germplasm core set for utilization in castor improvement. Ind Crops Prod 112:766–774

Anjani K, Reddy PAV (2002) Genetic divergence and potentiality of castor (*Ricinus communis* L.) germplasm collected from north–east India. Indian J Plant Genet Resour 15:157–159

Arslan M (2012) Evaluation of genetic similarity and agronomic traits of castor bean populations naturally grown in the Eastern Meditrranean region of Turkey. Biotechnol Biotechnol Equip 26:3089–3093

Bajay MM, Zucchi MI, Kiihl TA, Batista CE, Monteiro M, Pinheiro JB (2011) Development of a novel set of microsatellite markers for castor bean, *Ricinus communis* (Euphorbiaceae). Am J Bot 98(4):87–89

Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, Melake-Berhan A, Jones KM, Redman J, Chen G, Cahoon EB, Gedil M, Stanke M, Haas BJ, Wortman JR, Fraser-Liggett CM, Ravel J, Rabinowicz PD (2010) Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotechnol 28(9):951–956

Chavan AA, Benko PS, Sakhare SB (2012) Genetic divergence in castor. Bioinfolet 9:784–786

Collard BCY, Mackill DJ (2009) Start codon targeted (SCoT) polymorphism: a simple, novel DNA marker technique for generating gene-targeted markers in plants. Plant Mol Biol Rep 27(1):86. https://doi.org/10.1007/s11105-008-0060-5

Cooper DN, Smith BA, Cooke HJ, Niemann S, Schmidtke J (1985) An estimate of unique DNA sequence heterozygosity in the human genome. Hum Genet 69 (3):201–205

Dong H, Wang C, Li W, Yang GX, Yang H, Wang YR, Chen MH, Li FJ, Feng Y, Chen G (2012) Castor germplasm diversity analysis using AP-PCR and RMAPD. Acad Period Farm Proc 2012–06.www.cnki.com.cn

FAOSTAT (2016). http://www.fao.org/faostat/en/. Accessed on Apr 2018

Foster JT, Allan GJ, Chan AP, Rabinowicz PD, Ravel J, Jackson PJ, Keim P (2010) Single nucleotide polymorphisms for assessing genetic diversity in castor bean (*Ricinus communis*). BMC Plant Biol 10:13. https://doi.org/10.1186/1471-2229-10-13

Gajera BB, Kumar N, Singh AS, Punvar BS, Ravikiran R, Subhash N, Jadeja GC (2010) Assessment of genetic diversity in castor (*Ricinus communis* L.) using RAPD and ISSR markers. Ind Crops Prod 32 (3):491–498

Gilbert JE, Lewis RV, Wilkinson MJ, Caligari PDS (1999) Developing an appropriate strategy to assess genetic variability in plant germplasm collections. Theor Appl Genet 98(6):1125–1131

Goodarzi F, Darvishzadeh R, Hassani A (2015) Genetic analysis of castor (*Ricinus communis* L.) using ISSR markers. J Plant Mol Breed 3:18–34

Goodarzi F, Darvishzadeh R, Hassani A, Hassanzaeh A (2012) Study on genetic variation in Indian castor bean (*Ricinus communis* L.) accessions using multivariate statistical technologies. J Med Plants Res 6:1160–1167

He S, Xu W, Li F, Wang Y, Liu A (2017) Intraspecific DNA methylation polymorphism in the non-edible oilseed plant castor bean. Plant Divers 39(5):300–307

Hu J, Vick BA (2003) Target region amplification polymorphism: a novel marker technique for plant genotyping. Plant Mol Biol Rep 21(3):289–294

Joshi CP, Zhou H, Huang X, Chiang VL (1997) Context sequences of translation initiation codon in plants. Plant Mol Biol 35(6):993–1001

Kahl G (2001) The dictionary of gene technology. Wiley-VCH, Weintiem

Kanti M, Anjani K, Usha Kiran B, Vivekananda K (2015) Agro-morphological and molecular diversity in castor (*Ricinus communis* L.) germplasm collected from Andaman and Nicobar Islands, India. Czech J Genet Plant Breed 51:96–109

Kanti M, Anjani K, Venkat Ramya K (2014) Molecular diversity in castor germplasm collection originated from north-eastern hill province of India. Int J Res Sci Inn 1–6

Lakhani HN, Patel SV, Bodar NP, Golakiya BA (2015) RAPD analysis of genetic diversity of castor bean (*Ricinus communis* L.). Int J Curr Microbiol Appl Sci 4:696–703

Lan XY, Chen H, Zhang YD, Sun WB, Lei CZ (2006) A new method of molecular marker—RMAPD. Hereditas 28(1):78–84

Law JA, Jacobson SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11(3):204–220

Li F, Wang C, He D, Liu Y, Chen M, Wang Y, Li F, Yang Z, Chen G (2012) Evaluation of genetic diversity in castor (Ricinus communis L.) using RAPD markers. Adv Mater Res 343:981–987

Li G, Quiros CF (2001) Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in Brassica. Theor Appl Genet 103 (2):455–461

Liu S, Yin X, Lu J, Liu C, Bi C, Zhu H, Shi Y, Zhang D, Wen D, Zheng J, Cui Y, Li W (2016) The first genetic linkage map of *Ricinus communis* L. based on genome-SSR markers. Ind Crops Prod 89:103–108

Lu Z, Qi JM, Fang PP, Su JG, Xu JT, Tao AF (2010) Genetic diversity and phylogenetic relationship of castor germplasm as revealed by SRAP analysis. J Wuhan Bot Res 28(1):1–6

Mei-Lian T, Yan MF, Wang L, Wang L, Yan XC (2012) Analysis of genetic diversity in castor bean by SRAP markers. J Mong Uni Natl 2012–05. www.cnki.com.cn

Onkarappa T, Sunil Kumar SV, Lakshmipathaiah OR, Vasanth KR (2014) Assessment of genetic diversity among castor genotypes (*Ricinus communis* L.) Res Environ Life Sci 7:135–336

Pranavi B, Sitaram G, Yamini KN, Dinesh Kumar V (2011) Development of EST-SSR markers in castor bean (*Ricinus communis* L.) and their utilization for genetic purity testing of hybrids. Genome 54 (8):684–691

Qiu L, Yang C, Tian B, Yang JB, Liu A (2010) Exploiting EST databases for the development and characterization of EST-SSR markers in castor bean (*Ricinus communis* L.). BMC Plant Biol 10: 278. https://doi.org/10.1186/1471-2229-10-278

Quintero V, Anaya-López JL, Núñez-Colín CA, Zamarripa-Colmenero A, Montes-García N, Solís-Bonilla JL, Aguilar-Rangel MR (2013) Assessing the genetic diversity of castor bean from Chiapas, México using SSR and AFLP markers. Ind Crops Prod 41:134–143

Raji AA, Anderson JV, Kolade OA, Ugwu CD, Dixon AGO, Ivan (2009) Gene-based microsatellites for cassava (Manihot essculanta Crantz): prevalence, polymorphism, and cross-taxa utility. BMC Plant Biol 9:118. https://doi.org/10.1186/1471-2229-9-118

Ramana Rao PV, Shankar VG, Pallani JVP, Rajesh V, Reddy AV (2012) Genetic diversity analysis for drought tolerance in castor using EST-SSR markers. Crop Res 44(1):152–155

Ramesh M, Lavanya C, Brahmeswar Rao MV (2012) Genetic divergence in some indigenous and exotic germplasm lines of castor (*Ricinus communis* L.) under rainfed conditions. Indian J Dryland Agric Res Dev 27:79–83

Ramesh T, Siva sankar A, Sree laxmi J, Durgarani CV, Gouri Shankar V, Gouthami P, Leela C, Narayana Reddy P (2014) Genetic divergence studies for yield components and root traits in castor (*Ricinus communis* L.). Indian J Plant Genet Resour 276:271–276

Ramesh T, Siva Sankar A, Sreelakshmi J, Palle G, Leela C, Durga Rani CV, Gouri Shankar V, Lavanya B, Narayana Reddy P, Dudhe MY (2016) Utilization of *in silico* EST-SSR markers for diversity studies in castor (*Ricinus communis* L.). Physiol Mol Biol Plant 22 (4):535–545

Rao VR, Hodgkin T (2002) Genetic diversity and conservation and utilization of plant genetic resources. Plant Cell Tissue Org Cult 68(1):1–19

Reddy KP, Nadigatla VPRGR, Mulpuri S (2015) Molecular diversity in castor (*Ricinus communis* L.). Ind Crops Prod 66:271–281

Rivarola M, Foster JT, Chan AP, Williams AL, Rice DW, Liu X, Melake-Berhan A, Huot Creasy H, Puiu D, Rosovitz MJ, Khouri HM, Beckstrom-Sternberg SM, Allan GJ, Keim P, Ravel J, Rabinowicz PD (2011) Castor bean organelle genome sequencing and worldwide genetic diversity analysis. PLoS One 6(7): e21743. https://doi.org/10.1371/journal.pone.0021743

Rukhsar Patel MP, Parmar DJ, Kalola AD, Kumar S (2017) Morphological and molecular diversity patterns in castor germplasm accessions. Ind Crops Prod 97:316–323

Rukhsar, Patel MP, Parmar DJ, Kumar S (2018) Genetic variability, character association and genetic divergence studies in castor (*Ricinus communis* L.). Ann Agrarian Sci. https://doi.org/10.1016/j.aasci.2018.02.004

Sarada C, Anjani K (2013) Establishment of castor core collection utilizing self-organizing mapping (SOM) networks. J Indian Soc Agric Stat 67(1):71–78

Sawant SV, Singh PK, Gupta SK, Madnala R, Tuli R (1999) Conserved nucleotide sequences in highly expressed genes in plants. J Genet 78(2):123–131

Senthilvel S, Shaik M, Anjani K, Shaw RK, Kumari P, Sarada C, Kiran BU (2017) Genetic variability and population structure in a collection of inbred lines derived from a core germplasm of castor. J Plant Biochem Biotechnol 26(1):27–34

Seo KI, Lee GA, Ma KH, Hyun DY, Park YJ, Jung JW, Lee SY, Gwag JG, Kim CK, Lee MC (2011) Isolation and characterization of 28 polymorphic SSR loci from castor bean (*Ricinus communis* L.). J Crop Sci Biotechnol 14 (2):97–103

Sharma A, Chauhan RS (2011) Repertoire of SSRs in the castor bean genome and their utilization in genetic diversity analysis in *Jatropha curcas*. Comp Funct Genomics 2011:286089. https://doi.org/10.1155/2011/286089

Simoes KS, Silva SA, Machado EL, Silva MS (2017) Genetic divergence in elite castor bean lineages based on TRAP markers. Genet Mol Res 16 (3). https://doi.org/10.4238/gmr16039776

Tan M, Xue J, Wang L, Huang J, Fu C, Yan X (2016) Transcriptomic analysis for different sex types of *Ricinus communis* L. during development from apical buds to inflorescences by digital gene expression profiling. Front Plant Sci 6:1208. https://doi.org/10.3389/fpls.2015.01208

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucl Acids Res 17(16):6463–6471

Tomar RS, Parakhia MV, Rathod VM, Thakkar JR, Padhiyar SM, Thummar VD, Dalal H, Kothari VV, Kheni J, Dhingani RM, Sabara P, Golakiya BA (2017) Molecular mapping and identification of QTLs responsible for charcoal rot resistance in castor (*Ricinus communis* L.). Ind Crops Prod 95:184–190

Udoh OE, Abu NE (2016) Phenotypic variability in Nigerian castor (*Ricinus communis* L.) accession. Afr J Agric Res 11:4222–4232

Vasconcelos S, De Souza A, Gusmao CLS, Milani M, Ana MBI, Ana CBV (2010) Heterochromatin and rDNA 5S and 45S sites as reliable cytogenetic markers for castor bean (*Ricinus communis,* Euphorbiaceae). Micron 41:746–753

Vasconcelos S, Onofre AVC, Milani M, Benko-Iseppon AM, Brasileiro-Vidal AC (2016) Accessing genetic diversity levels of Brazilian genotypes of castor with AFLP and ISSR markers. Pesq Agropec Pernamb 21:24–31

Vinay PS, Shanker goud I, Prabhakaran AJ (2017) Genetic diversity assessment among advanced breeding lines of castor (*Ricinus communis* L.). J Pharmacog Phytochem 1:610–613

Vivodik M, Balazova Z, Galova Z (2014) RAPD analysis of genetic diversity of castor bean. Int J Biol Biomol Agric Food Biotechnol Eng 8, 7

Vivodik M, Balazova Z, Galous Z (2016) Genetic diversity analysis of castor (*Ricinus communis* L.) using SSR Markers. J Microbiol Biotechnol Food Sci. https://doi.org/10.15414/jmbfs.2016.6.2.777-780

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M (1995) AFLP: a new technique for DNA fingerprinting. Nucl Acids Res 23(21):4407–4414

Wang YR, Li FJ, Liu Q, Yang Z, Li W, Wang L (2013a) Study on analysis of genetic relationship of 37 main castor varieties by peroxidase isozme. North Hort 2 (02):55656

Wang C, Li G, Zhang ZY, Peng M, Shang Y, Lyo R, Chen Y (2013b) Genetic diversity of castor bean (*Ricinus communis* L.) in Northeast China revealed by ISSR Markers. Biochem Syst Ecol 51:301–307

Wang ML, Dzievit M, Chen Z, Morris JB, Norris JE, Barkley NA, Tonnis B, Pederson GA, Yu J (2017) Genetic diversity and population structure of castor (*Ricinus communis* L.) germplasm within the US collection assessed with EST-SSR markers. Genome 60 (3):193–200

Yadav P, Anjani K (2017) Assessment of variation in castor genetic resources for oil characteristics. J Am Oil Chem Soc 94(4):611–617

3 Classical Genetics, Cytogenetics, and Traditional Breeding in Castor Bean

C. Lavanya, A. Vishnuvardhan Reddy, Bhramar Dutta and Rajib Bandopadhyay

Abstract

Castor being a perennial, cross-pollinated, sexually polymorphic crop with high environmental sensitivity, initial plant breeding efforts were restricted to plant height and duration. Domestication of a wild, perennial crop to an annual crop of medium plant height and duration is the first success. Further, development of a two-line breeding system and standard seed production technology led to successful commercial exploitation of heterosis. Being a monotypic genus, diversification of parental base is restricted to intra-generic, intraspecific, or inter-varietal hybridization. Phenotypic expression is highly plastic and varies with locations and seasons. Majority of the morphological characters are monogenic, independently assorted with very limited linkages among the traits. However, information on genetics of major morphological characters is scattered in several old publications. Conventional breeding methods were successful in developing about 40 high-yielding hybrids and varieties with inbuilt resistance to major pests and diseases. An effort is made in the present chapter to consolidate the information on genetics and breeding methods followed in India and elsewhere.

3.1 Introduction

The castor plant, a member of Euphorbiaceae or spurge family, is a commercially valuable, non-edible oilseed crop due to its high yield potential even under optimum management conditions. It is an ideal commercial crop due to its high oil content (45–55%) of seed with high levels of a unique fatty acid, ricinoleic acid (80–90%). Ricinoleic acid (12-hydroxyl-*cis*-9-octadecenoic acid) has more than 250 industrial uses in addition to biofuels. Castor is cultivated in more than 29 countries over an area of 14.48 lakh ha with a production of 19.48 lakh tons and productivity of 1346 kg/ha (2015–16). India has the largest area of 8.3 lakh ha with a production of 14.2 lakh tons and 1713 kg/ha productivity. Though indigenous to Eastern Africa and Ethiopia, it is reported as polyphyletic in origin with four major centers of diversity, viz. Iran–Afghanistan–(former) USSR region, Palestine–South West Asia, India–China, and the Arabian Peninsula.

Castor, a drought-tolerant crop, usually grows well in relatively dry and warm regions having a well-distributed rainfall of 500–750 mm. In

C. Lavanya (✉) · A. Vishnuvardhan Reddy
ICAR-Indian Institute of Oilseeds Research, Hyderabad 500030, Telangana, India
e-mail: c.lavanya@icar.gov.in

B. Dutta · R. Bandopadhyay
Department of Botany, UGC-CAS, The University of Burdwan, Golapbag, Burdwan 713104, West Bengal, India

C. Kole and P. Rabinowicz (eds.), *The Castor Bean Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97280-0_3

heavy rainfall areas, the crop puts an excessive vegetative growth and assumes a perennial habit. Castor requires a moderate temperature (20–26 °C) with low humidity throughout the growing season to produce maximum yields. A *kharif* crop is also being cultivated as a *rabi* crop in southern states like Telangana, Andhra Pradesh, Tamil Nadu, and Odisha with assured irrigation. The crop can withstand long dry spells but cannot establish under waterlogging conditions.

3.2 Monotypic Genus

Ricinus is a monotypic genus, and the species *communis* is subdivided to six subspecies, viz. *persicus, chinensis, zanzibarinus, sanguineus, Africans,* and *Mexicans* based on eco-geographical grouping (Kulkarni and Ramanamurthy 1977; Moshkin 1986; Weiss 2000). The species *R. communis* was also considered as a composite species (sp. *collectiva*) with three elementary species, while many considered this species as a monotype. There is no difference in the chromosome number ($2n = 20$) among the subspecies, and they all can cross easily with each other (Kulkarni and Ramanamurthy 1977; Atsmon 1989). Genetic variability though restricted to intraspecific variability, variable phenotypic expression due to its cross-pollinated nature and independent assortment of several phenotypic characters provided a wide genetic base for the breeders.

3.3 Cytogenetical Studies

The somatic diploid chromosome number of *R. communis* is $2n = 20$. The karyotype analysis indicated that chromosomes differ in size (4.6–19.2 μM) and distribution of heterochromatin. Castor is considered to be a secondary balanced autopolyploid with $N = 10$ (Richharia 1937). Based on the maximum number of secondary associations (5) observed at I and II metaphases, the haploid constitution was postulated as AAA, BB, CC, DD, E. Polyploidy has been induced using colchicine (0.3%) on apical meristems (Narain 1953). Increase in stomatal size and decrease in stomatal density were observed in colchicine-induced tetraploids of three castor genotypes, viz. 48-1, DCS-107, and AP-41 (ICAR-IIOR 2017). Pollen fertility varied from 0 to 35% in 48-1 and AP-41 mutants, while quadrivalent associations were more frequent in the tetraploid mutants.

3.3.1 Idiogram

The idiogram is constructed in different species of *Ricinus* to correlate the chromosome morphology variation with respect to changes in sex expression. Jakob (1956) has shown that each ten chromosome pairs of the compliment is easily morphologically distinguishable in pachytene stage of meiosis cell division, by the regions which stained deeply with acetic orcein. "Chromatic" region for deeply stained zones was coined by Brown in 1949 in *Lycopersicum esculentum.* Macro-chromosome-bearing region of pachytene chromosome of *Ricinus* sp. apparently supports the view. The micro-chromomere of the chromatic zone is not clearly distinguishable, whereas the achromatic zone, flanking both sides of the chromatic zone, shows the presence of lightly stained micro-chromomere. The ten pairs of chromosomes were numbered by Paris et al. (1978) in numeric fashion, whereas Alexandrov and Karlov (2016) marked the chromosomes pairs in alphabetic order. Paris et al. in (1978) described the details of chromosome morphology. It is clear from Fig. 3.1 that Chromosome 1 is the longest and least heterochromatic chromosome of the complement. Chromosome 2 is the main nucleolar organizer (NOR). Chromosome 3 contains proximal heterochromatic region. Chromosome 4 is second longest chromosome after Chromosome 1. Chromosome 5 is metacentric as it has two arms almost equal in length. In Chromosome 6, centromere is flanked by one large chromomere in long arm and proximal chromomere is in the one-third way of centromere. A NOR region is present in Chromosome 7 as of Chromosome 2.

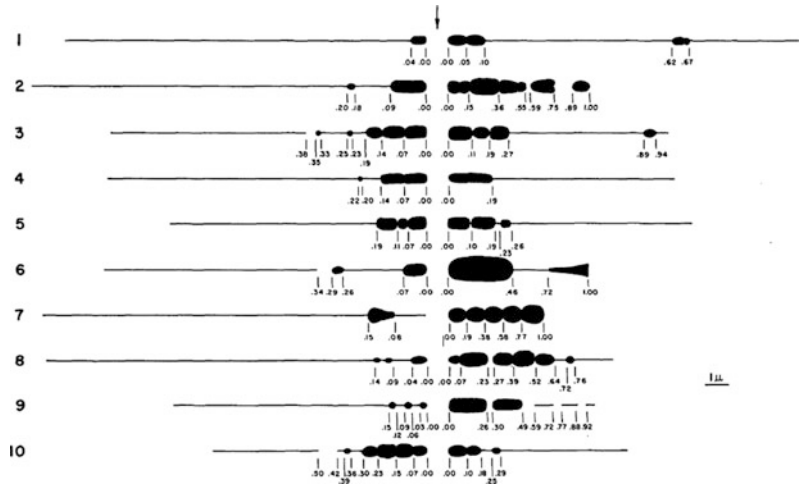

Fig. 3.1 Idiogram of ten pachytene bivalents of *Ricinus communis*; where centromere represented by arrows, long arms are in the left side and short arms are in the right side, euchromatic regions are the thin line and thickened lines represent heterochromatic regions. The space between the lines represents secondary constriction. Adapted from: Paris et al. (1978)

However, the NOR in chromosome 9, previously reported by Jelenkovic and Harrington in 1973, was not confirmed. Chromosome 8 is more or less similar with Chromosome 4. Chromosomes 9 and 10 are the shortest.

3.3.2 Molecular Cytogenetics of Castor Bean

The study of molecular cytogenetics spread some light on genome sequencing of castor plant. The organization of the major DNA repeat sequences can be studied by fluorescence in situ hybridization (FISH) mapping. Earlier, several authors (Paris et al. 1978; Zhong et al. 1996) characterized pachytene chromosome by high-resolution FISH technique. But FISH-based mapping done for the first time showed that 45S rDNA and 5S rDNA signals on the pachytene chromosome are on the short arm; contrarily, Vasconcelos et al. (2010) showed that they were localized on the long arm. Alexandrov et al. in 2016 experimented the FISH-mediated karyotype which constitutes ten pairs of chromosomes shown in Fig. 3.2. Six chromosomes among them, viz. B, C, D, F, H and J, emit heterochromatin bands and rcsat39 probe labeled with digoxigenin-11-dUTP signals that are depicted in Fig. 3.3. A chromosome pair which is non-heterochromatic carries rcsat390 probe signals. The telomeric probe rcsat47 emits faint yellow signal, if it is placed at the chromosome terminals.

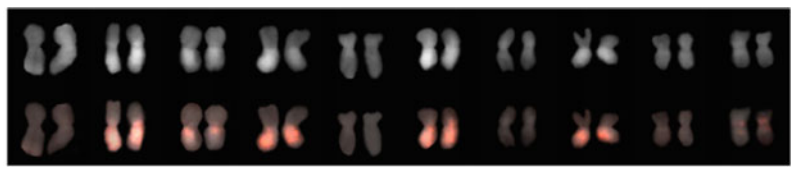

Fig. 3.2 Castor bean Karyotype based on FISH signals. Adapted from: Alexandrov and Karlov (2016)

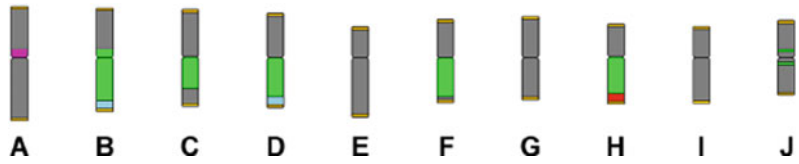

Fig. 3.3 FISH signal-based ideogram of castor bean where 45S rDNA represent in cyan color, 5S rDNA in red, rcsat39 probe emits green, magenta emitted by rcsat390 probe, telomere produces very little yellow signals. *Source* Alexandrov and Karlov (2016)

3.4 Pollination Mechanism

Due to monoecious nature, castor is highly cross-pollinated, and wind is the main agent for pollination indicating cross-pollination to the extent of 5–36% (Sesadhadri and Muhammad 1951). Pollen can travel up to even 1 km on a clear day with normal wind velocity, but no pollen movement happens on a rainy day. Weather factors like continuous cloudy weather during flowering, low temperature due to continuous rains cause infertility as low temperature, and/or high humidity produce abnormal pollen grains and also restrict the free movement of viable and healthy pollen.

Castor is a cross-pollinating plant but unlike other cross-pollinating crops, it is inclined toward self-pollination with low inbreeding depression (Moshkin 1986). Flowers of central spike are inclined toward self-pollination (30–70%), and flowers of lateral spikes pollinate by cross-pollination. In large areas of the crop, pollen flies up to 2.5–3 km. Maximum quantity of pollen was on the level of castor plants (60 cm), and it was ten times less at a height of 3 m. Isolation distance varied from as low as 300 m for certified seed production to 1500 m for nucleus seed production of female lines (Zaveri et al. 2010; Varaprasad and Lavanya 2015). Due to its low inbreeding depression, breeding techniques are not restricted to that of a cross-pollinated crop.

3.5 Genetics of Major Morphological Characters

Three morphological characters including stem color, bloom, and spines on the capsules are essential for characterization or description of a genotype. Several other characters like plant type, leaf shape, nature of the spike, echinate nature of stem, and petiole are also useful as morphological descriptors.

3.5.1 Stem Color

Stem color is a stable, distinct morphological trait of monogenic inheritance. A wide variation ranging from bright green to dark red through green with reddish–bluish tinge on sunny side, carmine or rose red, mahogany red, sulfur white, etc., is observed in the germplasm. Earlier botanists classified stem color into five categories while classification as three categories, viz. red, green, and mahogany, or two categories as among the cultivated varieties, green and faint red is most common.

Stem color is due to the pigmentation of sap color in epidermal palisade cells and parenchymatous areas of the stem and used to classify the genotypes (Kulkarni and Ramanamurthy 1977). Recent studies conducted under Distinctiveness, Uniformity, and Stability (DUS) tests on anthocyanin pigmentation starting from seedling to

flowering stage indicated that green is the basic, distinct color while the presence of anthocyanin pigmentation over the green color turns to red, while its magnitude and intensity result into either light red, mahogany, or dark red. The intensity of anthocyanin pigmentation varies with sunshine, presence and intensity of bloom, and age of the plant, etc. Stem color is more deep and intense when there is no bloom on any plant parts or zero bloom. In DUS testing, stem color is determined with the color charts of Royal Horticultural Society, after the removal of bloom as it distorts the original color (Chakrabarty et al. 2006, 2009).

Genetic studies conducted during 1920 s indicated monogenic inheritance for stem color and dominance of colored stem over green stem. In crosses between red-blush × green, Kulkarni (1959) reported monogenic inheritance where F_1 was all red-blush with 3 red-blush: 1 green in F_2 which was later confirmed by Solanki and Joshi (2001), Rao et al. (2005), Lavanya and Gopinath (2008), and Gourishankar et al. (2010). Epistatic ratios for stem color indicated two factors M and G for three colors, where factor G converts tinged green to green and mahogany to rose and M is the factor for mahogany (Seshadri and Mohammd 1951). A single-gene *Rst/rst* for stem color in three crosses of castor also indicated digenic epistatic ratio of 9 green: 6 red: 1 mahogany where green stem color is dominant to red stem color (Lavanya and Gopinath 2008). The presence of either of two genes *Rst* or *Gst* individually produces red stem color, while the presence of both the two genes *RstGst* polymerizes the effect to turn to green color. The presence of two recessive genes *rstgst* produces mahogany stem color in cross DPC 13 × M 574. Mahogany color has both green and red colors with predominance of red color.

A distinct dark purple color morphotype, collected from Assam and Manipur states, however showed maternal inheritance in a cross between purple and green color morphotypes (Anjani et al. 2007).

3.5.2 Bloom

Bloom or waxy coating on the stem and other parts like leaves, capsules, and spines in castor serves as a natural protection against extremes of weather and infestation of insect-pests (Seshadri and Seshu 1956; Chandrasekharan and Sundaraj 1946). Cold injury and leaf hopper incidence are higher in plants without bloom than in plants with bloom, while it is vice versa for white flies. The intensity of ashy white bloom increases with sunlight and decreases with the fog or humidity. Castor plants are classified mainly based on the absence or presence of bloom on all external plant parts as zero bloom or with bloom. However, further classification is based on the presence of bloom on combination of plant parts, as single bloom (stem + petiole + capsule stalks), double bloom (stem, petiole, capsule + lower side of the leaf), and triple bloom (all the above parts + upper side of the leaf) (Kulkarni 1959; Narian 1961; Kulkarni and Ramanamurthy 1977).

Bloom is monogenic dominant or partially dominant over zero bloom (Kulkarni 1959; Kulkarni and Ramanamurthy 1977; Lavanya and Gopinath 2008). Single bloom was monogenic and dominant over zero bloom while double bloom had a dominant digenic complementary action of 9 double:3 single:4 zero bloom in a cross between double bloom and zero bloom (Peat 1926). Double bloom was controlled by two complementary genes B and C, where B alone expresses single bloom while C can express double bloom only in the presence of B (Peat 1926). The variation in the intensity of bloom was also controlled by another dominant gene D. Triple bloom was always dominant to other bloom variations (Seshadri and Muhamed 1951; Narian 1961; Pathak and Asthana 1962). Dominant nature of a single gene for bloom (PtBl) was confirmed in F_2 ratios of 3 triple bloom: 1 single and double bloom (Lavanya and Gopinath 2008).

The intensity of bloom within the plants varies with the age of the leaves; mild in the youngest to traces in matured, senescing leaves but highest in the physiologically active leaves. Even a series of single bloom families were heterozygous for the intensity of bloom varying from normal to heavy bloom. In triple or double bloom types, the presence of bloom on upper or lower side of the leaf is mostly confined to the latest emerged leaf indicating the role of penetrance and expressivity of the genes controlling the bloom character or due to multiple alleles.

3.5.3 Plant Type

The concept of plant type in castor gained importance especially with Texas stable pistillate-tenth-order revertant (TSP-10R), an exotic pistillate line introduced from USA. It was further crossed with three other inbred lines, viz. JI-15, JP-5, and 26006 at Vijapur, Gujarat in 1968 to develop an indigenous, stable pistillate Vijapur Pistillate-1 (VP-1), with distinct characters like dwarf stem, condensed nodes, cup-shaped leaves, convergent branching, long primary spike, and non-shattering character. Majority of the pistillate lines derived from VP-1 background have dwarf stem, condensed nodes, cup-shaped leaves, and convergent branching which are highly linked characters (Rao et al. 2005). The plant type of VP-1 was an instant success due to its feasibility in hybrid seed production for agronomic management and wind pollination. In addition, plant type is monogenic inheritance where dwarf plant type is recessive to the normal plant type with elongated nodes, normal plant height, and divergent branching (Moshkin 1986; Lavanya and Gopinath 2008; Gourishankar et al. 2010). Single dominant gene (*PtN*) is reported for normal, elongated plant type segregated in F_2 as 3 normal (N): 1 dwarf (Dw) in two crosses between Normal × Dwarf and Dwarf × Normal while in another Dwarf × Normal cross, F_2 segregated to fit into 45:19 ratio indicating the role of three genes for plant type (Lavanya and Gopinath 2008).

The distance between stem color (*Rst/gst*) to plant type (*PtDw*) and single gene for pistillate character and plant type was estimated as 40.2 and 47.6 centimorgans based on square root of frequency of double recessive phenotypes in F_2 data (Lavanya and Gopinath 2008).

3.5.4 Leaves

Leaves in castor are simple, small to large size and palmate with 7–11 lobes. The genotypes differ in the size and number of the leaves which is also influenced by the environmental factors like season, soil type, and water stress. Papaya leaf type, with deeply dissected leaf lobes and serrated margins, is a distinct trait, observed in germplasm and exotic collections. The trait developed in parental lines like DPC-12, DPC-15, DCS-12, DCS-59 is of monogenic, recessive control and serves as a distinct morphological marker. The petioles of castor leaves are long (30 cm to > 40 cm) restricting the adoption of high plant density. A short petiole plant (< 20 cm), was used to transfer the monogenic trait to a high-yielding cultivar FCA-PB.

3.5.5 Number of Nodes to Primary Spike

Node number is an indication of flowering initiation, and on an average, every node takes 4–5 days to develop (Shifriss 1964; Moshkin 1986). Node number varies between the locations and planting season, while it is constant for each genotype within the locations based on planting season (Lavanya and Gopinath 2008). Genotypes were classified as early (< 12 nods), medium (13–16 nodes), and late (> 17 nodes) based on the number of nodes to the primary spike. Low node number is dominant to higher node number and segregates as 9 early:6 medium:1 late types. Digenic or trigenic epistatic ratios controlled the character and varied with the parents of the crosses involved (Lavanya and Gopinath 2008).

3.5.6 Spike

3.5.6.1 Spike Compactness

In castor, spikes are classified based on the arrangement or density of capsules on the spike as loose, compact, and semi-compact (Kulkarni and Ramanamurthy 1977). Compact and semi-compact spikes are highly susceptible to fungal diseases like Botrytis gray mold under conditions of high humidity, rainfall, and cloudy weather due to poor aeration and ventilation. In local collections, the basal portion of the raceme up to 15–20 cm did not bear any capsules, while in improved varieties like HC-1 and HC-6, the first portion (of 13 cm) of raceme bore 70% of the total capsules (Kulkarni 1959).

Inheritance pattern of spike nature indicated monogenic inheritance, either incomplete dominance or dominance of the compact versus loose spike (Solanki and Joshi 2001; Lavanya and Gopinath 2008). The distinction between compact and semi-compact is not much clear while loose spike is distinct from compact or semi-compact spikes. Dominance of a single gene, *SpSc,* for compactness of the spike was indicated in 3 semi-compact: 1 loose spike in a cross between compact spike type (viz. DPC 13) and semi-compact (viz. M 574). In another cross between two pistillate lines of compact vs. semi-compact, two epistatic inhibitory genes controlled the spike type while F_2 population fit into the ratio of 13 semi-compact: 3 compact spike types. The presence of dominant gene at one locus [*SpSc*] and recessive gene at other locus [*Spll*] produced the same phenotype—semi-compact spike while presence of dominant gene at other locus [*spL*] resulted in loose spike types.

3.5.7 Capsule Color

Capsules with purple, mahogany, sulfur white and green colors were available in the germplasm. The green color of capsule was controlled by a single dominant gene (Patwardhan 1931; Sesharadri and Muhammad 1951). Similarly in a cross between green and sulfur white, the F_2 is segregated in the ratio of 15 normal green:1 sulfur white.

3.5.7.1 Capsule Spines

Spines on capsules in castor are a misnomer as they are not of pricking type. Spines vary in length, number, and distribution on capsules. The number of spines may vary from 0 to 150 per capsule. Based on the presence and distribution of spines on capsules, genotypes were classified as spiny, non-spiny, and semi-spiny. In some instances, capsules on main raceme were non-spiny while those on secondary and tertiary were spiny indicating xenia effect (Pathak et al. 1965). Capsules with long and densely packed spines as in DCH-519 provide congenial micro-climate like high humidity, retention of raindrops, moisture for the germination of spores of Botrytis gray mold. Capsules with short spines, sparsely distributed spines, or absence of spines as in, viz., DPC-9, GCH-4, and 48-1 indicated tolerance to Botrytis gray mold. The presence of spines is partially dominant to non-spininess and controlled by a single gene, while F_2 gave a good fit to the 1 spiny:2 partial spiny:1 non-spiny capsules indicating incomplete dominance of sparse spiny nature and monogenic control of spiny capsule (Narain 1961; Anjani 1997; Chandramohan 2002). In a cross between spiny (130 spines per capsule) and non-spiny, the F_1 was intermediate (68 spines) and F_2 segregated in the ratio of 1 spiny:2 intermediate:1 non-spiny (Sehardri and Muhammad 1951; Kulkarni 1959). The non-spiny capsules have either rugged, wrinkled, or smooth surface. The rugged surface of capsule was partially dominant over smooth surface and controlled by a single dominant gene. Further, a single dominant gene controlled the stalk of the capsule over sessile capsule. The stalk of the capsule may be branched; partially dominant over non-branched and F_2 segregated in the ratio of 1:2:1 indicating monogenic inheritance of capsule stalk.

3.5.7.2 Locules

Normally, the capsules are trilocular but sometimes bi- and tetralocular capsules were also noticed. Rarely, capsule with 4–8 locules were recorded (Rao and Thandavarayan 1954; Pathak et al. 1965).

3.5.8 Fruit Dehiscence

Capsules of all local varieties are shattering type, dehiscing at maturity compelling premature harvesting leading to low oil content. Breeding efforts led to non-shattering varieties where threshing became difficult. Partial dehiscence of capsules is preferred where in membranes covering the seed remains intact even though capsules dehisce preventing the shattering of the seeds on the ground (Kulkarny and Ramanamurthy 1977).

3.5.9 Seed Color

Seed color at the base varied from light chocolate, deep chocolate, red, purple, black, and white and is further modified by mottling patterns, which are of independent inheritance. Seed coat color is a distinct, stable morphological marker along with either low or high mottling pattern. Genotypes were classified based on the size of the caruncle as small, viz. GCH-2 or big, viz. DCH-177 (Chakrabarty et al. 2009).

3.5.10 Cotyledons

Being a dicot, castor has two cotyledons. A single cotyledon was observed in Cimarron variety at Berhampur (Savithri and Aiyadurai 1961; Savithri 1963).

3.6 Floral Morphology

Castor is an indeterminately determinate plant as each branch is terminated by a raceme or spike which develops in a sequential order. Unisexual flowers occur in clusters of three or more, spirally on the axis. Male flowers are protected by fused tepals, split into 3 or 5 segments, five stamens with branched, tree-like filaments, in branched clusters with distinct anther cells. Female flowers have spathaceous caducous calyx, a single superior three-celled ovary, short or long styles terminated by two or three bifid stigmatic branches, each of which is divided into two fleshy lobes with a papillose surface (Kulkarni and Ramanamurthy 1977).

The floral morphology of castor is described based on the arrangement of male and female flowers on the spike. Several classifications were given by different authors based on the number and pattern of arrangement of male and female flowers on the spike (Moshkin 1986; Kulkarni and Ramanamurthy 1977; Weiss 2000).

Castor, though a sexually polymorphic species, monoecious is still the most common floral morphology, with male flowers on the lower side while female flowers are on the upper half of the same raceme. Occasionally, both male and female flowers on either side were also observed (Joshi 1926). Normally, the male flowers occupy 30–50% of lower part of the raceme. At Kanpur, lines with 90–98% femaleness were recorded.

Based on the proportion of male and female flowers, the racemes are broadly classified as (1) mostly female: with one cluster of male flowers at the bottom of raceme; (2) partially female: upper half with female flowers and lower half with male flowers; and (3) mostly male: with female flowers on top one-third portion of the raceme. Plants either with 100% pistillate flowers (Joshi 1926) or with 100% male flowers were recorded. Presence of bisexual flowers was also reported (Joshi 1926; Seshadri and Muhammad 1951; Konwar 1960). Monoecious trait is monogenic and dominant over pistillate nature, but highly sensitive to environment and variations were also noticed with age and nutrition.

The basic sex forms in castor are monoecious, pistillate, and sex reversion either to monoecious or interspersed staminate flowers. **Monoecious (M)** spike has basal one-third to 1/2 male flowers while the top portion has female flowers. The

proportion of male flowers may vary usually from 30 to 50% of lower part of the raceme to 2–10% (Lavanya 2002). In between these few whorls have both male and female flowers in an interspersed fashion. However, in few cases, both upper and lower parts of the spike had both male and female flowers. Variations like topmost male flower, lowermost female, male flowers interspersed with female flowers at the tip of the spike are common. The extreme variation of highly male or mostly male occurs with 1–5 female flowers throughout the entire spike and in different orders.

Pistillate (P) occurs as a rare recessive mutant with the spike having female flowers throughout the central and lateral spike orders. Variations like mostly female or inclined toward female produce female flowers from the base to the entire length of the spike while few staminate flowers are clustered at the base of the peduncle and sometimes dispersed at the lower end only. **Interspersed Staminate Flower (ISF)** is a variant of pistillate form with male flowers interspersed, throughout the female flowers on the spike.

Sex revertant is a female that turns to monoecious at later stage. Shifriss (1960) and Moshkin (1986) classified the sex variants based on the proportion of female and male flowers on the spike.

3.6.1 Classifications of Sex Revertants or Variants

Shifriss (1960) classified sex variants in five categories; Type A-monoecious, Type C-pistillate, Type B-sex revertant from C to A, D to A, E to A, Type D with an apically interspersed raceme as a monoecious variant, Type E uniformly interspersed raceme as a female variant. Among the above variants, A and C are developmentally persistent forms.

Moshkin (1986) classified the sex variants as stable female (female flowers from primary and later spike orders), unstable female (primary pistillate while later orders partial or fully monoecious), inclined toward female (up to ten male flowers in the lower part of the spike), interspersed (few male flowers interspersed throughout the spike) and monoecious and male with a possible occurrence of individual hermaphrodite flowers in any of the forms (Fig. 3.4).

3.6.2 Sex Reversion

Sex reversion in castor or the genetic instability of female line is the major problem in hybrid seed production in castor. Shifriss (1960) in his extensive study on sex variants in castor

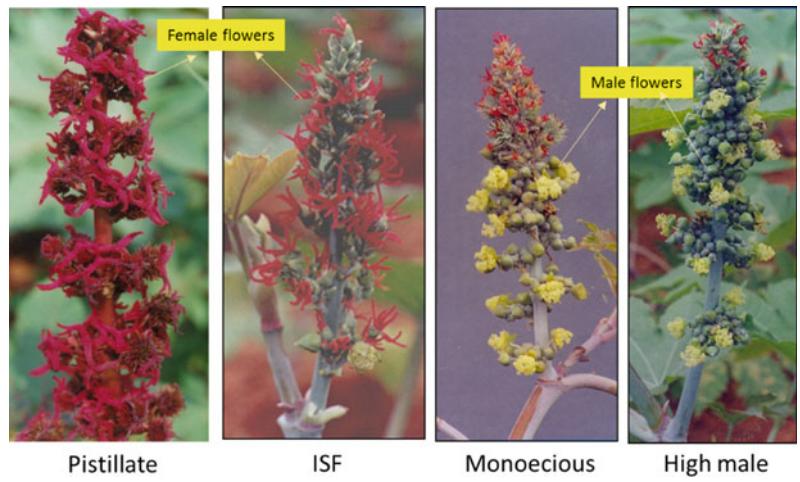

Fig. 3.4 Sex variants in castor

classified them as conventional and unconventional types. In conventional variants, qualitative genes determine the production of male or female flowers, while polygenes either accelerate or deplete a growth regulator to determine the male-to-female flower ratio (Shifriss 1960). Unconventional variants evolved as dominant spontaneous female mutants from monoecious inbreds are predominantly early-reverted. The genetic instability in all spontaneous female mutants was not linked with male sterility or meiotic abnormalities (Shifriss 1960). The time of phenotypic reversion is unstable, but reversion as such is not associated with any serious breakdown in the female producing mechanism. Reversions in later spike orders, seventh order or more, called late revertants were heritable and predictable to produce higher proportion of female plants, which in turn were late revertants. Development and maintenance of female lines in India commercially exploited the late revertants for production of quality hybrid seed while early revertants of less than third spike order are low heritable and unpredictable.

3.6.3 Inheritance of Sex Expression

Castor is typically monoecious plant, but the proportion of male and female flowers is greatly influenced by non-genetic factors. In normal monoecious varieties, the percentage of pistillate flowers on racemes is usually the highest on main raceme and decreases gradually on subsequent raceme orders. A proportionate increase in the number of male flowers is related to proportionate decrease of pistillate flowers and is highly influenced by seasons. The number of pistillate flowers and subsequent number of capsules on the main raceme have high heritability value (Sindagi 1964).

The expression of pistillate character is recessive to monoecious trait in majority of the cases and controlled by single to polygenes, depending on the material involved (Louis et al. 1986; Patel et al. 1986; Chauhan et al. 1992); additive and epistatic gene effects for the inheritance of 100% pistillate whorls in primary and later two orders (Solanki and Joshi 2000). Monogenic recessive gene control of perfect femaleness was indicated in M_2 of three female mutants of Aruna using 100-125 KR gamma rays (Weiss 1971). In 16 pistillate mutants, two pairs of closely linked recessive genes, *ppII* for pistillate expression and narrowly lanceolate functional leaves on floral axes and peduncles of inflorescences were obtained in Yong 283 (Zhou and Gu 1990).

3.6.4 Sex Expression Versus Environment

The role of environment for variation in the sex ratio of flowers within a genotype was well established and classified as female and male promoting environment. The tendency to produce female flowers in a spike increased with several environmental factors like winter season, low monthly mean temperatures (< 30 °C), young age or early spike orders, high nutrition, less difference between maximum and minimum temperatures, etc. On the contrary, summer or rainy season, high monthly mean temperatures (> 32 °C), old age or late spike orders, low nutrition, and large difference between maximum and minimum temperature promote male flowers on a spike (Lavanya 2002). The role of exogenous and endogenous growth hormones like gibberellic acid, silver nitrate, and ethylene in shifting the female and male tendency has been well documented (Ramesh et al. 2000; Lakshmamma et al. 2002, Murthy et al. 2003; Neeraja et al. 2010).

3.6.5 Inheritance of Interspersed Staminate Flowers (ISFs)

Two pairs of recessive genes, *id1* and *id2*, result in expression of ISF, while two, dominant genes, *Id1* and *Id2*, lead to the absence of ISF and are partially dominant to *id1id2* (George 1966). The highest number of ISF with homozygous recessive genes (*id1id1id2id2*) was produced irrespective of any environment. Higher peroxidase

activity was detected in male flower buds, which had five isoenzymatic bands than in female flower buds, while female lacked one and had two extra electrophoretic bands in peroxidase extracts from female flower buds (Jaiswal and Kumar 1983). Pistillate with ISF expression is recessive to pistillate trait and controlled from one (P_1) or two (P_1P_2) to four genes ($P_1P_2P_3P_4$). Digenic epistatic ratios like 13 pistillate: 3 pistillate with ISF in secondary spikes and tetragenic epistatic ratios in primary (162:94) and secondary (229:27) spikes were also observed in two crosses (Lavanya and Gopinath 2008).

3.7 Genetic Parameters

3.7.1 Correlations

Seed yield, a complex trait of low heritability, is essentially dependent on the expression and correlation of different component characters. Correlation studies help in simultaneous selection of positively associated characters, correlated response, and indirect selection. Several studies on correlation in castor indicated that seed yield was positively correlated to total and effective primary spike length, weight of the primary raceme, total number of capsules per primary or plant, total number of effective spikes per plant, 100-seed weight, number of nodes and plant height (Ramesh et al. 2001, 2010; Lavanya et al. 2006; Severino et al. 2012), plant height, number of leaves, branches, stem girth, petiole length, leaf area, number of capsules per plant and dry matter accumulation (Shinde et al. 1985), while oil yield is closely and positively associated with plant height, effective length of primary raceme, number of spikes, number of capsules and seed yield on whole plant basis (Ratnakumari 1996).

Under rainfed conditions, significant and positive correlation between seed yield and number of capsules per plant was reported while days to 50% flowering and number of nodes to primary spike had negative significant correlation with seed yield (Patel and Jaimini 1991). Days to flowering recorded positive correlation with plant height, number of nodes on main stem, and number of capsules on primary raceme, while it was negatively associated with number of secondary branches and total number of racemes per plant (Bhatt and Reddy 1981). Several other yield components had negative association among themselves like seed yield with duration and node number (Dorairaj et al. 1973a), number of spikes with total and effective length, capsule number (Ratnakar 1982), and capsule number with test weight (Shinde et al. 1985).

3.7.2 Path Coefficient Analysis

Majority of the characters like number of nodes to primary spike, total primary spike length, number of capsules on main spike, number of effective spikes per plant, and 100-seed weight had positive direct effect on seed yield (Lavanya et al. 2006; Omkarappa et al. 2010; Ramesh et al. 2010), while a few studies reported negative direct effect of number of nodes to primary raceme, plant height up to primary raceme, effective spike length, number of capsules on main spike, 100-seed weight, oil content on seed yield (Ananthasayana and Reddy 1981; Raju 1981; Patel and Jaimini 1991; Yadav et al. 2004).

Number of capsules on the primary spike had a positive indirect effect through other characters except 100-seed weight on capsule yield (Ananthasayana and Reddi 1981), while plant height, seed yield from primary spike, and number of lateral spikes indirectly increased yield per plant (Moshkin 1986). Majority of the traits like plant height, seed yield of primary spike, number of lateral spikes, days to flowering, number of nodes up to primary raceme, total and effective length of primary spike had an indirect effect on seed yield per plant (Ramesh et al. 2010), while oil content and 10-seed weight had a large negative direct effect on seed yield under irrigated conditions (Yadav and Raviprakash 2004).

3.7.3 Gene Action

Studies on predominant gene action and their heritability assist the plant breeders to formulate the immediate breeding methodology to be followed and further selection programs. Predominance of non-additive gene action for seed yield and many other yield components was reported by many researchers (Lavanya et al. 2003; Madariya et al. 2008; Barad et al. 2009; Patel et al. 2010). Additive gene action was predominant only for plant height, length of the raceme (Narkhede et al. 1987), and all the characters (Mehta et al. 1991). Both additive and non-additive gene actions were also reported for some of the characters (Mehta et al. 1991; Dobariya et al. 1992; Padhar et al. 2010; Ramesh et al. 2010).

Complementary epistasis was observed for days to flowering, effective branches, node number on main stem, and 100-seed weight in three different crosses, while for other characters, duplicate epistasis was recorded (Pathak et al. 1988). Tertiary racemes had the highest genetic advance under moderate stress conditions when ten genotypes with divergent plant type and duration were evaluated under early and late sown conditions. Breeding methods like mass selection for reducing the duration and improving the seed yield and biparental mating with reciprocal recurrent selection, bulk population, and mass pedigree methods of selection for improving the seed yield were suggested (AICRP on castor, DOR 2006).

3.8 Heritability

Among the several yield components, apart from morphological characters, 100-seed weight is a highly heritable trait, while other characters like total and effective primary spike length, total number of capsules per plant, number of effective spikes per plant, plant height, number of nodes to the primary raceme, days to 50% flowering and maturity are mostly influenced by local environmental conditions. However, higher narrow sense heritability estimates for maturity, flowering time, and low for seed yield were reported (Singh and Yadava 1981; Ratnakumari 1996).

Number of nodes had the highest heritability (Kaul and Prasad 1983), while all the characters except plant height showed higher heritability values (Dhapke et al. 1992). High heritability coupled with high genetic advance and genotypic variability was noted for number of branches/plant, number of capsules/main raceme, length of pistillate region, capsules/plant, and seeds/plant suggesting the influence of additive gene action (Dhapke et al. 1992).

In a study on genetic association and selection indices in three height groups of castor, plant height markedly affected trait associations and selection index (SI) of traits, including seed yield and it was more efficient than selection based on seed yield alone (Bhatt and Reddy 1986). The selection index for dwarf varieties included number of primary and secondary branches/plant and capsules/raceme, while for moderately tall varieties, number of primary branches/plant, days to maturity, and seed weight were important. Selection indexes for tall varieties were days to flowering, node number, and plant height. Seed yield and racemes/plant were included in the five-trait SI for all varieties.

Further, high heritability coupled with high genetic advance and genetic variability were found for branches/plant, capsules/plant, and seed/plant indicating the influence of additive gene action providing scope for further selection. Mass selection was effective for shortening the duration and improving the yields (Reddy et al. 1999). Estimates for heritability and genetic advance were higher for secondary over primary yield components like capsule number, weight, seed number, and 100-seed weight indicating a higher selective value for these characters.

Heritability ranged from 0.152 for seed yield/plant to 0.893 for number of nodes on main stem. Genetic advance was high for plant height and number of days to flowering. High positive phenotypic correlations were shown by all characters except for number of days to flowering and node number. Path coefficient analysis indicated that both the number of capsules/primary raceme

and number of secondary branches exhibited high positive direct effects on seed yield/plant. In yet another study, Bhatt and Reddy (1986) further suggested that semi-dwarf lines with a large number of capsules on primary raceme and with a moderate number of primary and secondary branches are desirable parents in a breeding program.

3.9 Worldwide History of Castor Plant Breeding

Castor, a native of Africa and India, is a very interesting crop with all possible transitions from an uncultivated plant to a weedy type, semi-cultivated to a field crop with no gap between uncultivated and cultivated forms. Under natural conditions, castor plants show a wide variety of growth habit and seed collection was done mainly for oil extraction (through crushing) at local markets. Production of high oil content seeds in a commercial scale was difficult mainly due to the natural height of the plant and prolonged flowering period.

Crop improvement of castor was initiated in the USA long back in early 1902, at the Brooklyn Botanical Gardens, New York and Oklahoma. Initial attempts were directed toward pure line selection for higher oil content from a collection of a large number of samples (White 1918b). Later in 1940 s, intensive work was carried out in 22 different states to combine different desirable characters like earliness, short plant, and non-shattering habit in different varieties like "Conver" and "Kansas" with oil content of 53%. Techniques for emasculation, hand pollination, and hybridization were standardized for effective breeding programs (White 1918a, b; Weibel and Woodworth 1946).

The breeding objectives in erstwhile USSR were aimed to develop genotypes with early maturity and non-shattering capsules. In the initial phase (1922–1931), the objective was to introduce varieties with early maturity to escape frost and uniform maturity suitable for manual harvesting. Breeders were successful to develop, Manchurian, an early maturing (75–85 days) variety. In the second phase (1932–39), efforts were directed toward introduction of non-shattering Sanguineus types suitable for double-stage combine harvesting. In the third phase (1939–1956), the aim was to introduce varieties suitable for single-stage combine harvesting (uniform/synchronous maturity) with simultaneous hulling of seeds. The emphasis is now shifted to development of high-yielding varieties with synchronous maturity, suitable for combine harvesting and resistant to harmful diseases such as Fusarium wilt. Among the several varieties developed, "Persian" variety with a single spike was the most productive variety (Moshkin 1986).

Intensive breeding efforts for desirable characters, specifically for only seed yield, high oil content, and spike length through different breeding techniques at State Agronomic Institute, Sao Paulo, EMBRAPA, Brazil, resulted in several high-yielding promising strains like No. 14, 38, and 45 (Kulkarni 1959).

In Italy, castor is an annual crop grown under irrigated conditions during spring–summer period, while recently the possibility of unirrigated semi-perennial crop is explored in eastern coast of Sicily (Anatasi et al. 2015). Initial breeding efforts through selection and hybridization based on germplasm collection led to early maturing (100 days), small plant height, like variety M-6. A Tunisian cultivar performed better for seed yield and ricinoleic acid (89%) compared to the local RG-2 (Sicilian genotype) (Anatasi et al. 2015).

Breeding efforts in France were carried out at the Central Plant Breeding Station, Versailles. Mutants generated from pollen grains by electromagnetic treatment had non-dehiscent capsules with thin walls.

In the Transvaal area of South Africa, near Johannesburg and Pretoria, Knapp worked on a large collection of geographically diverse germplasm from Italy, India, Manchuria, Africa, Brazil, and Russia for over 20 years. Screening for drought resistance under extreme hot summer and cold winter conditions of Transvaal (1800 m above sea level) resulted in annual varieties with drought resistance and tolerance to *Alternaria*

ricini. Other promising varieties of non-spiny, non-shattering, and short types were cultivated in several other African and Asian countries like India, Pakistan, and Iran.

Breeding work in India was initiated in early 1920 s at different places like Tindivanam, Rajendranagar, Hebbal, Raichur, Nagpur, Jalgaon, Nadiad, Junagadh, Jullandhar, and Kanpur. Crop improvement in castor was initiated as selections from populations with an emphasis on seed yield and branching habit. Later on, non-shattering character and oil content were given importance in the fifties. Selections from local populations resulted in only 10–20% improvement for seed yield and 1–2% for oil content for want of large stock of germplasm. In 1960 s, castor was the first oilseed crop to induce variability for duration, plant height, and oil content using ionizing radiations, viz. *X*- and gamma rays and chemical mutagens. Later on, with the introduction and development of pistillate lines, hybrid development took a momentum and up to 70% of the castor-growing areas are dominated by mostly public sector hybrids. The emphasis is now on development of high oil yielding, early and medium duration (90–150 days) hybrids, and varieties resistant to major pests and disease like wilt complex, Botrytis gray mold and sucking pests (Lavanya et al. 2006).

In the Philippines, the most common varieties (*tangan-tangan* in Pilipino) grown in the country are *Bangkok, Brazilian, Ethiopian, and Lamao Red* which differ in their seed color, plant height, and adaptability. Among the two types of Bangkok, Bangkok brown-spotted type is generally adapted to the Philippine conditions with 49–56% oil. Bangkok white-spotted type, on the other hand, has few small chestnut white spots scattered on its back side.

The *Brazilian* variety, commonly grown in Mindanao and Luzon, is tall (6–8 ft) with dark brown stem color and 49.3% oil content. Ethiopian is an early maturing, agronomically adaptable variety with red seed having small white dots on both the sides and 49% oil. Another variety is *Lamao Red* which is tall (6–8 ft) with reddish brown stem color. Other castor bean varieties include *Cimarron, Connex, Baker No. 1, Baker 195,* and *Iranian* variety (Anon. 2007).

3.10 Breeding Methodology

3.10.1 Selection from Germplasm or Introduction

Castor, a cross-pollinated crop with limited inbreeding depression, is often treated as a self-pollinated crop in breeding programs (Severino et al. 2012). Castor, being a monotypic genus, subspecies and local germplasm collection were the base material for selection program. At V.S. Pustovoit All-Union Scientific Research Institute (VNIIMK), USSR, high-yielding, non-shattering varieties like Donskaya 172/1 and Kruglik 5 were selected from Persian subspecies, while early maturing Saratovskaya 66 and Ceripi Wild were selected from *ruderalis* subspecies of African and Asian countries (Moshkin 1986). In USA, tall and late maturing *zanzibarian* subspecies were used for selecting early, short, non-shattering, and high oil content varieties like Cimarron, Baker 296, Lynn, Dawn, Hale, Campins with high yield, short internodes, and better foliage (Moshkin 1986; Hegde and Lavanya 2012). In India, selections from local germplasm led to selection of several medium to late maturing, medium plant height, non-shattering, high oil lines like HC-1 to HC-9 in India (Lavanya et al. 2006).

3.10.2 Mass Selection

Mass selection is an effective breeding method in case of heterogeneous local landraces with highly heritable qualitative and quantitative characters. In castor, it is used mainly for selection of female plants, long primary spike, and reduced plant height. Mass selection followed by progeny row selection was effective to select wilt-resistant plants and improvement of local selections (Moshkin 1986). Significant achievements by mass selection included Kavkazskaya

from VNIIIMK, USSR, IAC-38 from Brazil, Conver and Kansas from USA (Kulkarni and Ramanamurthy 1977).

A revised mass selection or family selection was also used to select high-yielding families and individual plants within the families. The method includes random-mating in isolation between new and unimproved old varieties followed by selection of the best plants along with parental types coupled with negative or reverse selection of low yielding, late and shattering plants. About 15–30 families and 100–150 individual plants within each family were selected. Mass selection along with progeny testing was effective for oil content and resistance to Fusarium wilt, viz. Fioletovaya, a variety improved for wilt resistance (Moshkin 1986). The method was also used to select early, short, annual types in place of perennial types.

Initial selection criteria like yield and branching habit were gradually replaced by other traits like non-shattering capsules and oil content (Kulkarni and Ramanamurthy 1977). Three varieties were purified for two variable characters like days to flowering and number of capsules after three cycles of mass selection (Reddy et al. 1999). A population CCP-1 was developed by inter-breeding with free flow of genes among selected lines of the varieties, viz. HC-6, HC-8, JI-35.

3.10.3 Single Plant Selection with Progeny Tests

The method was followed initially in germplasm collections having natural genetic variability for several heritable traits. Single plant selections for highly heritable traits followed by progeny row testing unravel the hidden, recessive genes. Several heritable traits can be simultaneously dealt by selecting highest possible number of self-pollinated lines. The method is similar to Pustovoit method of recurrent selection in sunflower and followed both after selfing and hybridization to reveal the hidden genotypic variability behind the phenotypic expression of heritable traits.

Half of the seed of the best plants within a family are sown and evaluated by progeny row testing method. Remnant seeds of the best plants within the progeny rows are sown in the next season and are allowed to cross-pollinate. Selection for specific characters like non-shattering capsules, plant height, sparse or less branching suitable for combine harvesting was practiced in VNIIMK, USSR (Moshkin 1986).

In open-pollinated populations, individual selections were made between and within families to maintain genetic diversity, while bulk selections between families are quantitatively restricted to 1–2% of the total number of plants. In the initial stages of selection, pistillate spikes which are prone to high cross-pollination are selected. Selection in segregating populations is initiated in F_2 but may be delayed up to F_4–F_6 when the homozygosity increases. In case of breeding for Fusarium wilt resistance, selections started in F_1 itself when incomplete dominance was involved.

Selection needs to be based on phenotype alone with minimum interference of environmental conditions. Precautions need to be taken for a leveled field with crop rotation and application of fertilizers to the preceding crop. Selection of plants is more effective under wider spacing (70 × 70 cm in USSR, 90 × 60 cm in India) for maximum phenotypic expression and yield potential of individual plants. Single seed per hill is most ideal for selection of plants with productive spikes, increased oil content, and low hull percent in seeds. However, the evaluation of varieties and hybrids needs to be done under good agronomic conditions with required fertilizer and irrigation conditions.

3.10.4 Development of Inbred or Pure Lines

Inbreeding or induced self-pollination is essential to develop homozygous and homogenous pure lines for morphological and quantitative characters. Due to minimum inbreeding depression on selfing, inbreeding was used to develop and maintain varieties or inbreds or pure lines.

At VNIIMK, USSR, inbreds for several traits like duration, seed yield, oil content, plant height, and morphological characters like stem color, intensity, and distribution of bloom and presence of spines were separated. Successful examples include red-stemmed inbreds from green Kruglik 5, non-spiny and tall inbreds from Sanguineus 401, purification of Kruglik 5 for plant height, spike compactness, and maturity through inbreeding (Moshkin 1986). Inbreeding was also essential for recessive characters like non-shattering capsules, shortened internodes, and single spike. Wilt resistance of varieties, viz. Chervonnaya, VNIIMK 165, improved; Fioletovaya and VNIIMK 360 improved by selfing wilt-resistant plants in wilt sick plot (Moshkin 1986). The percentage of female plants was also increased by selfing or sibbing plants with highly female spikes (Moshkin 1986; Lavanya et al. 2006).

Inbreeding was practiced either from germplasm or working collection or interspecific breeding material. Segregation for plant height, length of spike, arrangement of capsules, maturity, and other morphological characteristics in addition to transgressive segregation was high in F_2 and fixed in later generations by selfing. Majority of the inbred varieties in India were developed by selection from the existing types followed by sibbing until homozygous condition is retained and evaluated in trials as inbreds (Kulkarni and Ramanamurthy 1977; Lavanya et al. 2006).

3.10.5 Hand Pollination and Selfing

Parental lines are sown in a crossing block, and morphological off-types are removed from them. Spikes with majority of unopened female buds and a few buds with extended stigma of yellowish green or light red color are selected either for selfing or crossing. The spikes, measuring 8–10 cm, are tagged one day prior to the pollination after removing male flowers, if any and covered with butter paper cover. Male flowers, which have just opened, from male parent are collected in a petri dish early in the morning, and the burst out pollen is used to pollinate female spike early in the morning from 6 AM to 9 AM on alternate days. Pollination has to be repeated 5–6 times depending on the spike length and opening of female flowers. Pollen is applied by brush or rubbing the male flowers pollen on the stigma of the female flowers. Details of crossing like parents, dates of pollination need to be mentioned on a separate tag instead of butter paper cover, as it need to be changed, minimum 3–5 times during the pollination period.

Sibbing is done with a mixture of pollens from plants of the same line. Selfing of highly female spikes is done either by inducing or forced selfing. At the beginning of the flowering, opened or fertilized female flowers were removed and enclosed in butter paper covers. Nipping induces male flowers within the whorls of female flowers and contributes to forced or induced selfing.

3.11 Hybridization and Selection

3.11.1 Intra- and Inter-generic Hybridization

The six subspecies within the monotypic genus, with $2n = 20$, are inter-crossable without any barriers or meiotic abnormalities. Each subspecies has its own strengths and limitations, and thus, intra-generic hybridization was successful in the initial stages of breeding program. In USSR, *communis* (earlier *sanguineus*), Persian and Chinese subspecies, were used as the initial breeding material.

3.11.1.1 Subspecies *Communis* (Sanguineus)

Originated from South west Asia, widespread near Mediterranean Sea and American continent. Plants are characterized by specific traits like large seed size, drought resistance, high seed yield, oil content, and indehiscent capsules but late maturing and unsuitable for mechanized harvesting due to their excessive vegetative growth.

3.11.1.2 Subspecies *Persicus* (Persian)

Originated from Asia Minor region, requires less heat, faster growth during vegetative period but susceptible to Botrytis gray rot due to its dense spikes and Fusarium wilt. Plants are characterized by small seed, productive, uniform maturity with dehiscent capsules. Varieties of Persian subspecies like Donskaya 172/1, Kruglik 5, Kavkazskaya improved were introduced in USSR (Moshkin 1986).

3.11.1.3 Subspecies *Sinensis* (Chinese)

Developed in East Asia under high humidity, has specific traits like early ripening, indehiscent capsules, drought resistant but low yielding, low oil content. Plants have higher number of lateral spikes, loose and weakly attached capsules resulting in losses during harvesting.

3.11.1.4 Subspecies *Indicus* (Indian)

Originated from India, tropical, prolonged vegetative period, tall, high yielding, indehiscent capsules, resistant to drought. Two small-seeded forms, variety indicus and variety griseofollius, were used to develop medium maturing varieties like SA-1 and SA-2.

3.11.1.5 Subspecies *Zanzibarian* (Zanzibarinus)

Is also tropical with slow growth rate, high yielding under optimum moisture conditions, and occurs in three tall and late types.

3.11.1.6 Subspecies *Ruderalis* (Ruderal)

Poorly developed, late maturing, low productivity and includes common, wild cultivated varieties (var. ruderalis) also includes wild growing varieties of Ethiopia and Asia Minor (var. Spontaneous), shattering local types from Sudan, Egypt, Mediterranean coast and huge tall forms from Mexico, Colombia, and Guatemala.

3.11.2 Examples of Intragenic Hybridization

A small-seeded variety microspermus developed by intercrossing the large-seeded communis subspecies with small-seeded Persian subspecies at VNIIMK, USSR, was the base material to develop the female line, CNES-1 at USA, and k-1182 at France. Nebraska 145/4, the female line of green stemmed variety was further used to develop high-yielding varieties like Pacific-6 at USA (Moshkin 1986). Varieties like VNIIMK 165, Stepnaya 6, large-seeded Donskaya, were also interspecific derivatives with majority of Persian characteristics. VNIIMK 165 is double-cross-derivative involving four subspecies, Persian (Kruglik 5), Chinese, communis, and Indian subspecies, and has non-shattering capsules and single spike for mechanical harvesting. Similar hybridization between early maturing, non-shattering, and drought-resistant Chinese or *sinensis* subspecies and local collections at VNIIMK led to dove colored, early maturing varieties like early hybrid, kubanskaya 9. Other interspecific derivatives include American variety Cimarron, suitable for irrigated conditions and varieties with condensed internodes of 1–3 cm, like Baker 296, Lynn, Dawn, Hale, Campinas with higher foliage and rapid growth of dry matter and yield. Dwarf female line, TSP-10R, is another widely used pistillate parent for development of hybrids in USA, Brazil, India, with a potential of 25–30 q/ha under irrigated conditions.

Initial attempts on inter-generic hybridization with related genera in the Euphorbiaceae family like *Euphorbia lathyris* were not successful (Moshkin 1986). Recently, putative hybrids were developed by using castor and cassava hybridization, and molecular tools were standardized to prove the hybridity of the material (Gedil et al. 2009).

3.12 Handling of Segregating Generations

3.12.1 Pedigree Method of Selection

The method is most commonly used to select simultaneously several heritable and morphological traits after hybridization. The possibility of recombinants is usually high, when parents are

genetically diverse. Selection criteria and detailed description of each selected plant are maintained in the registers. In castor, three highly heritable characters, viz. stem color and presence, and distribution of bloom and spines on the capsules of individual selected plants are maintained in the registers. Promising selected hybrids were selfed to generate a large F_2 population. Individual desirable plants from F_2 s were selected, selfed, and advanced to F_3, F_4 to derive uniform lines based on highly heritable characters. Continuous selfing will lead to a high degree of homozygosity in F_5 and F_6 generations where selection is based on quantitative characters. In F_6, selfed seed of individual best plants with uniform characters within a family is bulked.

Several such inbred lines are evaluated in preliminary varietal trials in unreplicated augmented randomized block design with standard checks replicated after every ten entries in each block. Simultaneous screening for resistance to pests and diseases either under natural or artificial inoculation conditions resulted in several high-yielding varieties with pest or disease resistance.

Pedigree method was used to transfer polygenic controlled character like long raceme. A number of long-duration (240–270 days), tall plant-type varieties like HC-1 to HC-8, EB-16 A, S-20, Junagadh 1, Punjab castor 1, EB 31, Rosy, MC 1 were developed by hybridization and selection methods prior to inception of All India Coordinated Research Project (AICRP) on castor. High-yielding varieties, viz. SA-I, SA-II, TMV-1, TMV-2, TMV-3, TMV-4, and TMV-5 developed at Tindivanam center were recommended for Tamil Nadu (Lavanya et al. 2006).

Major drawback in pedigree method of selection is that generation of diversity is limited to the initial population size in F_2 which is essential to produce F_3 inbred lines. The F_2 population size is determined by the diversity of the parents, number, and heritability of the characters involved; greater the diversity, larger the F_2 population size.

3.12.2 Bulk Method

Segregating populations are allowed to open pollinate without any artificial selection until the later generations, F_5 or F_6 or F_7, attain homozygous and homogenous populations. The method is more successful for selecting segregating generations both under abiotic and biotic stress conditions like drought, salinity, acidity, and disease resistance followed by preliminary yield evaluations.

3.12.3 Backcross Method

This method is most successfully used to transfer monogenically inherited qualitative characters in an otherwise highly adaptable, high-yielding background. Fixation of the various desired characters in the breeding material allowed the use of backcross method to transfer simply inherited characteristics, viz. short internodes, non-spiny capsule, stem color, presence or absence of wax on the stem, seed shattering, plant height, and disease resistance.

3.13 Recurrent Selection

The efficiency of recurrent selection breeding procedure for reducing the plant height or altering the plant stature of Guarani cultivar for mechanical harvest was established after four cycles of recurrent selection in three locations at Brazil (Auld et al. 2009). In the first stage of selection cycle, short plants were selected and self-pollinated. In the second stage, 180 self-pollinated lines were evaluated for plant height in isolation and 30 plants were selfed (Filho 1999). The selected lines were gone through five cycles of selection, and the 30 selected lines were intercrossed and the seed bulked to generate the cycle 1 seed. The procedure was repeated for four additional cycles of selection, and reduction in plant height ranged from 3.4 to 28 cm for the five cycles of selection, while seed yield was not influenced by reduction in plant height (Oliveira and Zanotto 2008).

3.14 Population Improvement

The method was initiated as a random-mating population involving hybrids, F_2 populations, improved varieties, inbred lines in a population and sown in 0.1–0.2 ha area under isolation for 2–3 years. Negative and mass selection was done to remove the unwanted plants and bulk the productive, healthy plants for evaluation in preliminary and multi-location trials, viz. Sovkhoznaya and Stepnaya 6 varieties (Moshkin 1986).

Breeding programs aim to combine high seed yield and oil content and thus need to exploit both non-additive and additive gene actions simultaneously through parental mating and recurrent selection. A population improvement program was initiated with an objective to isolate high oil lines with 6–25 nodes to flower. Single crosses involving three high oil lines, CO-1 (58.6%), HO, HC 8 (56%) and 15 other diverse male lines, viz. Pb-1, T-3, T-4, 411, Baker, 4-4, Sowbhagya, 413 A, 1-21, Aruna, 239, 279, Bhagya, VI-9, JI-44 formed the base material. Part of the F_1 seed was used to raise the segregating population. Biparental mating was attempted in each segregating population by crossing 50–100 selected plants from promising crosses to broaden the genetic base. Promising recombinants of the crosses were crossed in F_3 generation to generate base material for further improvement (Lavanya et al. 2006). Similar efforts initiated in major All India Coordinated Research Project (AICRP) castor centers, through construction of gene pools for monoecious and pistillate trait to generate diverse parental material.

3.15 Mutation Breeding

Creation of variability in castor is limited to intra-genetic, inter-varietal hybridization in absence of wild species and related genera. Mutation breeding was resorted to induce variability for morphological characters, sex expression, and resistance to Fusarium wilt. Basic studies on relative efficiency of chemical and physical mutagens, standard doses, and genotypic variability of mutagen sensitivity were studied in detail during 1970 s (Kulkarni and Ramanamurthy 1977).

In castor, three types of radiations, viz. gamma rays 20 to 45 KR, fast neutrons 2.5 × 1012, 5 × 1012, 1 × 1013 and 5 × 1013, and thermal neutrons 0.87 × 1013, 1.75 × 1013, 2.62 × 1013, 3.49 × 1013 were used to induce variability. Irradiations of HC-1 by castor breeders resulted in several morphological variants, viz. chlorotic leaves, distorted ovaries, sex variants, and cytological abnormalities like inversions, translocations, anaphase bridges, and ring chromosomes. Other mutagens like X-rays (50–1100 R) resulted in variants for zero and single bloom from triple bloom. Irradiation of HC-6 with thermal neutron treatment of 0.87 × 1013 mh/cm^2 led to isolation of a short duration mutant, NPH-1 and later released as Aruna variety. The variety is early flowering (35-40 DAS), maturing (110-150 DAS), short height (75 cm) with 11 nodes to primary raceme. It was popular in several south Indian states like Andhra Pradesh, Tamil Nadu, and Karnataka under rainfed conditions.

Three female mutants obtained using 100-125 KR gamma rays confirmed monogenic and recessive nature of the pistillate trait (Chauhan et al. 1992). Other mutagens like gamma rays, ethidium bromide, diethyl sulfate were also utilized to induce variations for plant type, growth duration, female spikes, etc. Gamma ray irradiation at 55-60 KR gamma rays, of VP-1, a wilt susceptible stable S-type pistillate line led to several wilt-resistant pistillate lines like M-574 and M-619 (Lavanya et al. 2001) which were used in the development of hybrids like DCH-519 and YRCH-2. Recent example includes, DPC-23, a triple bloom, wilt- and leafhopper-resistant, early pistillate with less than ten nodes from DPC-9, a zero bloom, wilt-resistant but leafhopper susceptible pistillate line (Lavanya et al. 2008).

3.16 Varietal Breeding

In castor, monoecious lines or male lines were also used as varieties initially before the development and popularization of hybrids. Conventional breeding programs like mass or pure line selection, recombination breeding followed by pedigree and backcross method led to the release of several varieties in erstwhile USSR, USA, Brazil, India, and several other countries.

Selection criteria for varietal development in erstwhile USSR varied from non-shattering, early maturing, short plant height, high seed yielding, synchronous maturity suitable for combine harvesting. The variety VNIIMK 165 developed in 1952 at VNIIMK gained popularity for its high seed yield, suitable for combine harvesting, non-shattering capsules, and uniform maturity. Since 1957, the emphasis has shifted to development of wilt-resistant varieties resulting in Chervonnaya, VNIIMK 165 improved, Donskaya long-spiked, and Donskaya early. The major donor for wilt resistance was small-seeded Sanguineus. Three major groups of varieties were identified for their high seed yield and wilt resistance (Table 3.1, Moshkin 1986). Breeding efforts were also directed toward varieties with shortened internodes like Karlik heterozygotic (Moshkin 1986), Dawn, Lynn (Brigham 1967, 1993), dwarf female lines like CNES-1, TSP-10R in USA, and Frantsiya 301 M. The variety SA-2 has outyielded VNIIMK 165 by 3 q/ha under Russian conditions.

Varietal development in India intensified after the initiation of seven AICRP castor centers, mainly to reduce duration, plant height, non-shattering, high seed yield, and oil content. Available varieties like Aruna, JI-44, GCH 3 were evaluated in seven centers, and breeding program was initiated by recombination breeding between available material and local selections, viz. RC-7, RC-12, RC-8, L.448.57, HC, 413 A, T3, 6501, Kalpi 6, 6501, Tarai-4, and TMV 2, in individual centers and selected for seed yield and important yield components. Consistent plant breeding efforts led to the release of two high yielding vaieties in 1974. Bhagya (65), a cross of high oil line (HO) and MI 415, which is earlier and shorter than Aruna and Sowbhagya (157-B) from a double cross involving Aruna and a short mutant, Mauthner's dwarf. Sowbhagya was suitable for intercropping systems due to its convergent plant type and slightly longer duration than Aruna. RC 8, an induced mutant from RC 1188 (a tall plant type from Tamil Nadu) is a medium maturing (150–180 days) variety with 1200–2000 kg/ha yielding ability, was released in Karnataka. Other promising varieties like AKC 1 for Maharashtra, GAUC-1 or VI-9 for Gujarat (a selection from S 20) were released under State Varietal Release Committees. GAUC-1 variety has an average productivity of 1200–1500 kg/ha with an oil content of 46–47%. SKI 73 is another early maturing (150 days) variety released for all irrigated areas of the country as GC 2.

Intensive cultivation of improved varieties without crop rotation and/or cultivation of castor in poorly drained soils led to increased buildup of Fusarium wilt and susceptibility of all the above selections. The breeding objectives were modified to develop high-yielding varieties resistant to Fusarium wilt or wilt complex.

Natural and artificial screening of available germplasm, crossing between wilt-resistant

Table 3.1 Promising varieties in USSR (1974–1978)

S. No.	Group	Specific characters	Varieties
1.	I	High seed yield, uniform maturity, suitable for combine harvesting, resistance to Fusarium wilt	VNIIMK 360, 3218, 4280, Sizaya 7, Chervonnaya, VNIIMK 165 improved
2.	II	High yielding, moderately resistant to Fusarium wilt	Donskaya long-spiked, Donskaya 39/44, Donskaya early, Stepnaya 6, VNIIMK 165
3.	III	Average yield, highly susceptible to Fusarium wilt	Early hybrid and Kruglik 5

Source Moshkin (1986)

sources and agronomically superior and high-yielding genotypes led to development of wilt-resistant varieties like 48-1 and DCS 9 from diverse wilt-resistant sources (Mauthner's Dwarf and 240). Variety 48-1 (Jwala) is the male parent of GCH 4 hybrid and recommended for endemic areas like Karnataka. It has non-spiny capsules, medium maturing 110–120 days to first picking with very high branching potential. The variety DCS 9 (Jyothi), recommended for rainfed areas of peninsular India, is an early maturing (90–100 days to first picking), wilt-resistant variety with high yielding capacity (1200 kg/ha) under rainfed conditions. Another popular variety, Kranti (PCS 4), was released from Palem center for rainfed areas of Andhra Pradesh which has early maturity 90–100 days to first picking with good branching potential. Other varieties like Haritha (PCS 124), a wilt-resistant variety, and Kiran (PCS 136), a non-spiny variety from Palem center, were also recommended for rainfed areas of Andhra Pradesh. Some of the promising varieties along with their yield potential, suitable areas for cultivation, and specific features are given in Table 3.2.

3.17 Heterosis Breeding

Exploitation of heterosis in castor was initiated since 1960 s even before the development of pistillate lines. Due to its inbuilt self-pollination nature and limited inbreeding depression on selfing, it was considered that heterosis may not be as high as in other cross-pollinated crops but it varied from < 20 to > 100% over the standard checks (Lavanya et al. 2006). Heterosis was observed for non-economical characters like plant height in early seedling stages, leaf number, and leaf area index. Attempts to exploit hybrid vigor through monoecious lines were not successful due to laborious process of emasculation. Heterosis was high for seed yield followed by number of capsules on the main raceme and 100-seed weight. Heterosis and heterobeltiosis for seed yield per plant were due to heterosis for capsules on main raceme, length of pistillate region of main raceme, effective branches per plant, and seed yield of main raceme while heterosis for seed yield was associated with number of effective spikes per plant (Lavanya et al. 2006). However, heterosis for seed yield, as

Table 3.2 Promising castor varieties in India

Variety	Yield potential (kg/ha)	Oil content (%)	Recommended states/regions	Salient features/traits
Kiran (PCS 136)	1200–1500 (R)	51	Rainfed areas of Andhra Pradesh and also late sown *kharif* conditions with one or two irrigations.	Tolerant to Botrytis gray mold
Haritha (PCS-124)	1400–1600 (R)	49	Light soils of Southern Telangana, Rayalaseema, and Prakasam district.	Resistant to wilt
Jwala (48-1)	1000 (R) 1800 (I)	50	All castor-growing areas under both rainfed and irrigated	Resistant to Fusarium wilt; tolerant to Botrytis gray mold, salinity
GC-3	2340 (I)	49	Irrigated areas of Gujarat	Resistant to wilt
Chandra prabha			Uttar Pradesh	Suitable for intercropping
DCS-107	1500–1700	49	Identified for both rainfed and irrigated areas of the country	Resistant to wilt and tolerant to leafhopper
JI-273 (GC-3)	2340	49.6	Irrigated areas of Gujarat	Resistant to wilt, tolerant to Macrophomina root rot
Pragathi (PCS-262)	1500–2500	48	Rainfed areas of Telangana	High seed yield and wilt resistant

per Atsmon (1989), may be due to the highly female expression inherited from the dominant female nature of the S-type pistillate line. Genetic basis of heterosis of seed yield is due to the factors other than heterosis per se like the improved parental lines for spike density, highly female spikes, earliness, and short stature (Atsmon 1989).

Heterosis was mainly manifested in parental lines of contrasting morphological characters like dwarf plant type with condensed nodes, cup-shaped leaves in pistillate lines vs. normal tall plant type, elongated nodes, flat leaves in male lines (Lavanya et al. 2006). Per se performance and average heterosis in dwarf × tall crosses were higher to the parents involving moderately tall × tall and tall × dwarf crosses. Heterosis for seed yield was correlated with heterosis for main spike length and capsules/primary spike when one of the parents was tall. The first-generation hybrid between Kruglik 5 and Sanguineus gave 1400–1900 kg/ha higher seed yield than both the parents in USSR (Moshkin 1986). Heterosis has increased up to 18–23% when parents were of diverse origin and upon emasculation of pistillate lines. Heterosis breeding gained its momentum due to the identification of pistillate lines and suitable male lines.

3.17.1 Types of Pistillate Mechanism Used in Hybrid Development

3.17.1.1 N-Type

The N-type pistillate lines, viz. Nebraska 145-4, CNES-1, were initially used in USA for development of castor hybrids. The pistillate character in N-type, governed by a recessive sex-switching gene and maintained by sib-mating, is similar to conventional rare recessive female mutants proposed by Shifriss. It is similar to genetic male sterility system as the progeny from female plants segregate into 1 monoecious: 1 pistillate in hybrid seed production plots. Seed production is laborious as normal monoecious plants have to be rouged out before anthesis leading to low genetic purity and high cost of rouging. The female line is maintained by allowing 20–25% monoecious plants in foundation seed production plots. The progeny of spontaneous sibbing of pistillate plants of four varieties, TMV-2, S-20, D-3, and HC 6, which segregated in the ratio of 1 monoecious: 1 pistillate was grouped under the N-type. CNES-1, the N-type pistillate line, developed at Davis, California, was also used in the development of pistillate lines at AICRP (castor) centers.

3.17.1.2 S-Type

S-type pistillate line was obtained by selection within sex reversals at the Weizmann Institute, Israel, and governed by dominant and epistatic effects. Sex reversals are plant variants, which begin as female and revert to normal monoecism at any time after the first raceme and ten or more racemes when grown as perennials. These perennial plants were considered as females, if grown only as annuals. Sex reversion is ontogenetically irreversible and is variegated where a part of the plant may still be pistillate, while the other half is reverted to monoecious.

Selection for femaleness within sex reversals is carried out by growing the plants as perennials and choosing the late revertants. Selection within a graded series of unconventional dominant female variants led to the development of S-type pistillate lines. Selfed plants of the second and third orders of reversion yielded more number of pistillate plants than sibbed pistillate and first selfed order revertants (Gopi et al. 1996) while fourth-order revertants gave significantly more number of pistillate plants than the early revertants (Patel and Joshi 1972). Selfed plants of the tenth order of reversion yield nearly hundred percent pistillate plants in their progenies. The rate of reversion in a female progeny is decreased from 79.2% in first-order revertant to 12.2% in fourth-order revertant. The highest number of capsules and seed yield was obtained when the fourth-order revertant was crossed with the male line having 50% flowers. Depending on the genetic background, the selection pressure, and

number of cycles of selection, female plants varied from 60 to 95% in the progenies from sex revertants when grown as annuals.

The development of a stable pistillate line, TSP-10R (Texas stable pistillate tenth-order revertant) in 1962, with 100% pistillate raceme on a high percentage of plants brought the momentum to the castor hybrids. The exotic line was further utilized to develop a stable pistillate line VP-1 in Gujarat, India.

VP-1 was the first stable pistillate line developed in Gujarat, India, from the segregation of a double-cross between F_2 of JHB 48 (JP 5 × 26006) × JHB 67 (TSP-10R × 719/1) with distinct morphological characters like green stem, triple bloom, cup-shaped leaves, condensed nodes, long primary spike with spiny capsules. Majority of the commercial hybrids were generated from S-type pistillate lines, viz. LRES-17, DPC-9, SKP-84, MCP-1-1, M-574, and M-619.

3.17.1.3 NES Type

In NES system, pistillate character is governed by a homozygous recessive gene, while interspersed staminate flowers (ISF) were manipulated by environmentally sensitive genes, but not confined to any particular raceme order (Ankineedu and Rao 1973). Temperatures of above 31 °C promote ISF, while lower temperatures result in fully female racemes. A pistillate line "240" of NES background was developed by a cross between non-revertant pistillate progeny of the cross 625-4 × HC 6 with US3 415-9. The male parent, US3 415-9, is an "E"-type sex variant with ISF from base to top. The pistillate population has 56% strictly pollinate and 44% pistillate with few ISFs that mature later than the topmost female flower appearing on the spike. Later, few pistillate lines like NES-6, NES-19, NES-22, JP-65 were developed in India.

3.17.2 Pistillate Lines

Prior to the development of pistillate lines, inbred lines having highly female spikes were used as female lines. Initial research by Sidorov and Sokolov led to the identification of a heterozygous female line K-57, which segregated into 1:1 female: male plants (Moshkin 1986). In USA, plants with partial or total femaleness were selected from the collection of Gilmore. Classen and Hoffman (1950) developed an N-type pistillate line Nebraska 145-4. A three-way inter-line hybrid was also developed using the female plants of the hybrid between N 145-4 and Brazil 330 with a male line.

The NES system of pistillate line 240 was developed by a cross between non-revertant pistillate progeny of the cross 625-4 × HC 6 with US_3 415-9. The male parent, US_3 415-9, is an "E"-type sex variant with ISF from base to top. The pistillate population has 56% strictly pistillate and 44% pistillate with a few ISFs that mature later than the top most female flower appearing on the spike (Kulkarni and Ankineedu 1966). Yet, another NES type from segregating generations of the cross (high oil mutant from HC 6 × Mauthner's dwarf) × (Aruna × Mauthner's dwarf) with desirable attributes such as low node number on main stem (12), normal plant type, medium tall, red stem, double bloom, non-spiny, and non-shattering habit with an oil content of 55% was identified (Ankineedu and Rao 1973). Temperatures above 31 °C promote ISF, while lower temperatures (< 31 °C) result in fully female racemes. The NES system is advantageous due to easy transfer and maintenance of a single recessive gene for femaleness in contrast to polygenic, dominant, and epistatic gene complex in S-type. JP-65, a NES-type pistillate line, was utilized for the development and release of a high-yielding hybrid, GCH-6 at Junagadh. Several other sources were identified by screening of 1250 germplasm accessions at ICAR-Indian Institute of Oilseed Research (formerly DOR, Directorate of oilseed Research) Hyderabad, viz. EC 153132, EC 169761, EC 169803 for pistillate character and used in hybridization program (1986–87). Four pistillate lines NES 15, NES 16, NES 17, and NES 18 were derivatives of these pistillate selections.

Diversification of pistillate base through hybridization, selection, and generation advancement for early-to-medium duration, stable pistillate character, wilt resistance, and

economic characters was taken up at ICAR-IIOR in collaboration with AICRP castor centers at SK Nagar, Junagadh, Mandor, and Palem. At Hyderabad, seeds of VP-1 were subjected to irradiation (55 KR gamma rays) and selections stable pistillate behavior and wilt resistance by screening in wilt sick fields five mutant VP-1 selections, viz. M-574, M-619, M-568, M-571, and M-591, were stable for pistillate behavior and resistance to wilt (Lavanya et al. 2000). Among the 20 pistillate lines developed at ICAR-IIOR, M-619, DPC 9, and DPC-11 showed resistance to wilt. In addition, pistillate lines NES 6, DPC 9, and 10 were stable for pistillate nature up to sixth order of spike irrespective of seasonal conditions. Genotypic variation for production of ISF was highest in NES-22, DPC-9, DPC-10, and M-568 during May to June. Good combiners for seed yield, 100-seed weight, oil content (DPC-12), number of effective spikes per plant (NES-6, DPC-11, and M-568) were identified (Rao et al. 2000). Stringent selection pressure for the maintenance of non-reversal and ISF nature of pistillate lines were the key factors for pistillate line stability (Chakrabarthy and Banu 1999). DPC-9, the most promising pistillate line with high femaleness, late revertant nature, and high seed yield (279 g/pl), was the female line of several hybrids like DCH-177, PCH-111, YRCH-1, and HCH-6. Clustering pattern indicated five groups of pistillate lines.

Majority of the pistillate lines are developed from VP-1 or S-type of stable pistillate line source, which is originated from an exotic source –TSP-10R from USA. This has been converted to suitable agronomic types of Indian conditions like LRES 17, DPC 9, DPC 10, DPC 14, DPC-15, DPC-16, DPC-17 to DPC-29 through conventional breeding techniques like intraspecific or inter-varietal hybridization followed by pedigree selection and mutation breeding.

A diversified source of pistillate character, i.e., 240, was introduced to local genotypes like Bhagya, CO-1 and selected for very early, dwarf, normal plant, good branching pistillate line called NES 6. The early nature of pistillate line resulted in very early hybrids with low seed yield and thus converted to a new agronomic background like TMV-5, using pedigree method of selection and a new pistillate line-DPC 16 with purple stem, zero bloom, spiny capsules with medium duration (12–14 nodes), unlike NES 6 (> 10 nodes) having a unique trait like hermaphrodite flower at the tip of the pistillate line was developed. A list of sources along with the pedigree and the pistillate lines developed in India is given in Table 3.3.

DPC-18, a medium duration, non-VP-1 based pistillate line, was generated by inter-varietal hybridization involving a triple-cross-derivative 163-1-10-2 (Bhagya × CO-1 × HC-8) and a wilt-resistant variety 48-1. IPC-23 (DPC-23), a mutant pistillate line developed from a zero bloom, leaf hopper susceptible DPC-9 pistillate line, with green stem, triple bloom, low node number (7–8), short plant height (40–50 cm), early flowering (30–40 days to 50% flowering), was resistant to leaf hopper and Fusarium wilt and good combiner for early flowering and maturity (Lavanya et al. 2008, 2018). Inter-varietal hybridization involving different castor hybrids, pistillate lines helped to generate

Table 3.3 Diversified sources of pistillate character in castor (ICAR-IIOR, Hyderabad)

Source	Pedigree	Pistillate lines developed
S-type	TSP-10R × VP-1	LRES-17, DPC-9, DPC-10, DPC-14, DPC-21, DPC-25
NES type	240	NES 6, DPC 15, DPC 16, JP-65
	163-1-11 × 1501-4	DPC 11
Mutant VP-1	VP-1 with 55 Kr gamma rays	M 571, M 574, M 619, M 584
Hybridization	163-1-10-2 × 48-1	DPC 18 (new)
	M 619 × JI 225 F_4	DPC 17 (new)

Table 3.4 Pistillate lines registered by PGRC, ICAR-NBPGR for unique traits

Pistillate line	Registration No.	Pedigree	Special features
LRES 17	INGR 01010	VP-1 × HC 8	Green, triple bloom, spiny, dwarf, condensed nodes, cup leaves, resistant to Jassids
DPC 9	INGR 01009	87-V-2-1	Non-revertant, S-type pistillate, green, zero bloom, spiny, resistant to Fusarium wilt
M 619	INGR 03095	Mutant of VP-1 (55 Kr gamma rays)	Green, triple bloom, spiny, dwarf, condensed nodes, cup leaves, resistant to Fusarium wilt
MCP 1-1	INGR 04121	Selection from VP-1	Mahogany, triple bloom, spiny, dwarf, cup leaves, resistant to leaf hoppers
DPC-16	INGR 14003	NES-6 × TMV-5	Pistillate line unique for hermaphrodite flower at the tip of the spike, purple stem, zero bloom

new pistillate sources through recombinants or transgressive segregates and about 15 diverse pistillate lines were stabilized after 12–15 generations of selection pressure for pistillate character. Table 3.4 is a list of five pistillate lines registered for their unique or specific traits by Plant Germplasm Registration committee (PGRC), ICAR-NBPGR.

Evaluation of pistillate lines in different sowing seasons indicated variation in the expressivity of the ISF character with either one or two male flowers or > 10 male flowers per spike in all the pistillate lines irrespective of the sowing season. Variation is due to a complex phenomenon of both genetic and environmental factors like monthly mean day temperatures at the time of spike initiation, triggering mechanisms or the role of modifying genes. Sowing in the months of July and August leads to high ISF production in female lines and low genetic purity of hybrid seed produced. September and October are ideal months for sowing of certified hybrid seed production while January is ideal for sowing of foundation seed production of pistillate lines.

3.17.3 Male Lines

Male or inbred line in castor is a monoecious line with a balanced proportion of both male and female flowers. The criteria for selection of an inbred line as a male line for testing its combining ability are based on several characters like earliness compared to the female parent, desirable agronomic characters, closely maturing secondary and third-order spikes, resistance to Fusarium wilt, non-shattering, balanced proportion of male and female flowers on the raceme, high 100-seed weight (28–35 g), and high oil content (48–55%). A dominant marker gene to estimate the hybridity in the F_1 especially a seedling character like stem or hypocotyl color is useful in early detection of rogues or off-types before flowering (Atsmon 1989). The ideal combination would be contrasting stem color characters like red and green for the male and female parents (Lavanya and Solanki 2010).

Several inbred lines were developed by intraspecific hybridization involving pest- and disease-resistant germplasm accessions in combination with agronomically suitable cultivars, parental lines, etc., either in single-, double-, triple-, or backcross followed by pedigree method of selection. Male lines with early or medium duration (150 days), short-to-medium height, high oil content, resistance to wilt complex, and insect-pests were evaluated for their combining ability in a series of line × tester crosses. Among more than 1000 inbred lines available in the country, some of the inbred lines like DCS-5, DCS-9, 48-1, DCS-78, DCS-89, DCS-107, TMV-5, and TMV-6 were repeatedly used in the hybrid development programs. Male lines with good agronomic adaptability, per se

performance, and combining ability are ideal for male lines (Costa et al. 2006; Lavanya and Varaprasad 2012).

3.18 Hybrids

The first castor hybrid in the world, GCH 3 (TSP-10R × JI-15), was an instant success due to its high yielding ability (88% yield increase over S-20), drought resistance, medium maturity (140–210 days), high oil content (46.6%) and was popular even in the rainfed castor-growing area. The problem of non-shattering was overcome in another early maturing hybrid, GAUCH-1 (VP-1 × VI-9) with 16% yield increase over GCH 3 and drought escape mechanism due to efficient root system than the varieties under receding moisture conditions (Reddy et al. 1999). Another hybrid GCH 2 (VP-1 × JI 35) was released in 1985 for irrigated areas of Gujarat with tolerance to root rot, 13% yield increase over GAUCH 1. The hybrid has spikes with interspersed male flowers increasing the number of capsules in higher-order spikes (Lavanya et al. 2006).

Among the 18 hybrids released so far in the public sector system, GCH-4 is high yielding (1200–2200 kg/ha), suitable for rainfed and irrigated conditions, tolerant to wilt and still the most popular hybrid even after 32 years of release. It also gave the world's highest seed yield (9 tons/ha), as a perennial crop under intensive cultivation with high inputs near riverbanks of Khisurpuri regions of Ahmedabad district (Lavanya et al. 2006). It was replaced by the latest high-yielding hybrid GCH-7 (3000 kg/ha), which is resistant to both Fusarium wilt and reniform nematode complex. Table 3.5 represents three early duration, high-yielding, wilt- and leaf hopper-resistant hybrids, viz. DCH 32, DCH 177, DCH 519, suitable for rainfed and irrigated conditions which were developed from ICAR-IIOR, Hyderabad. Hybrids like YRCH-1, PCH-111, PCH-222 were released for rainfed conditions of Tamil Nadu and Telangana states of Southern India. More than 70% area of castor is under hybrid cultivation with a seed replacement rate (SRR) of 100%. However, for remaining area SRR is about 28–30% with varietal cultivation. Castor, being a commercial crop in Gujarat, about 43 private seed companies registered 88 experimental hybrids for commercial sale during 2010 (agri.gujarat.gov.in). More than 95% of the castor-growing area in Gujarat is occupied by castor hybrids, and the rise in productivity is spectacular from 350 to 1970 kg/ha (Damodaram and Hegde 2010).

3.19 Resistance Breeding

3.19.1 Resistance to Fusarium Wilt

Fusarium wilt caused by *Fusarium oxysporum* f. sp. *ricini* (*F.o.ricini*), a soil-borne fungus, is a destructive disease causing 39–77% loss in production (Pushpavati et al. 1998). The success of any resistance breeding program relies upon the basic information on etiology of pathogen, inheritance pattern of resistance, and development of standard screening procedures, availability of resistant sources, and reliable breeding methods for incorporation of the disease resistance (Lavanya and Solanki 2010). Standard screening procedures through development of sick plot and artificial inoculation in pot culture conditions were developed in AICRP castor system (Santha et al. 2014). A large number of germplasm accessions, breeding lines, varieties, and hybrids were identified in sick plots developed at three AICRP castor centers (Table 3.6). Several inbred lines were developed by intraspecific hybridization, involving wilt-resistant sources either in single-, double-, triple-, or backcross followed by pedigree method of selection.

The inheritance and genetics of resistance to Fusarium wilt varied with each cross and controlled either by recessive genes (Sviridov 1988; Podkuichenko 1989; Lavanya et al. 2011; Rao et al. 2005), dominant genes (Reddy et al. 2010, 2011), interaction of duplicate genes (Sviridov 1988); polygenes (Desai et al. 2001), or oligogenes with epistatic interactions (Lavanya et al.

Table 3.5 Salient features of castor hybrids in seed chain

Hybrid	Mean seed yield (kg/ha)	Areas recommended	Salient features
GCH 4	1200 (R) 2200 (I)	Both rainfed and irrigate areas, all over country	Resistant to leafhoppers, tolerant to Fusarium wilt
GCH 5	1800 (R) 2800 (I)	Rainfed and irrigated areas of Gujarat	Red, double bloom, medium duration (120–180 days), semi-spiny, wilt tolerant.
DCH 177	1550 (R) 2130 (I)	Rainfed areas of Andhra Pradesh, Karnataka, Tamil Nadu, Maharashtra, and Orissa	Red single bloom, spiny, early duration (90–150 days), resistant to Fusarium wilt, both parents are resistant to wilt
DCH 519	1740 (R) 2130 (I)	All over the country	Green, triple bloom, spiny, resistant to Fusarium wilt, leaf hoppers, and both parents are resistant to wilt.
GCH 7	3000 (I)	Irrigated areas of Gujarat	Resistant to nematode-wilt complex
HCH-6	1800 (R)	Karnataka	Resistant to wilt and white fly
GNCH-1	2500–3000 (I)	Late kharif, rabi Gujarat	Resistant to wilt and leaf hopper
GCH-8	1895 (R) 3590 (I)	All over the country	Resistant to wilt and leaf hopper
YRCH-2	2100 (R)	Tamil Nadu	Resistant to wilt

R Rainfed; *I* Irrigated

Table 3.6 Parental lines and hybrids resistant to wilt complex

Disease	Male lines	Pistillate lines	Hybrids
Fusarium wilt	SKI 232, 271, 291, 293, 252, 263, JI-273, 274, 296, 298, 299, 303, 314, 315, 319, 320, 322, 326, 327, 331, 338, 340, 342; DCS 5, 9, 33, 57, 84, 85, 86, 86-1, 89, 97, 98, 100, 102, 103, 104; PCS 124, 171, M-36-03, 35-03, 3-04 (PVT 16-03), 7-04 (F6-235-3-03).	SKP 23, 42, 72, 84, 106, 112, 117, 118, 119, Geeta, JP 81, 83, 85, 86, 88, 90, 93, 96, DPC 9, 11, M 584, 571, 619	GCH 4, 5, 7, DCH 177, 519
Macrophomina root rot	JI-220	JP-81, 86, 88, 89, 93 and 96, M-584, 619	GCH 6
Reniform nematode	–	M-619	GCH-7

2011). However, for the development of a wilt-resistant castor hybrid, both the parents should be wilt resistant (Desai et al. 2001; Lavanya et al. 2011). The inheritance of resistance in 48-1, a highly resistant variety is governed by two recessive genes (Rao et al. 2005). In a cross between susceptible and resistant parents, hybrids have a tendency toward susceptible parent (Golakia et al. 2005). Breakdown of resistant cultivars is a serious problem, especially in high-intensive cultivation without proper crop rotation. Heterotic and wilt-resistant hybrids should be further exploited through recurrent selection and inter se mating in segregating generations for developing wilt-resistant parental lines (Patel and Pathak 2011).

Several sources of resistance to Fusarium wilt were incorporated in high-yielding, agronomic background leading to the development of several inbreds, pistillate lines, and hybrids (Lavanya

et al. 2006). Prominent among them were DCS-9, 48-1, Haritha, DCS-107, DPC-9, M-574, M-619, DCH-177, DCH-519, GCH-7, etc.

Several preliminary hybrids, advanced breeding material, parental lines were identified as wilt-resistant consecutively for two years while DCS-86, DCS-105, DCS-107, DCS-118, DPC-23, and M-571 were the ideal wilt-resistant parental lines with consistent wilt reaction (9–18%) compared to susceptible check, JI-35 (94–96%) in the sick plot (Santha et al. 2016).

3.20 Summary

The present chapter on classical genetics and traditional plant breeding methods is an attempt to consolidate the scattered information available on different aspects. A brief history of genetics of different morphological characters and the complexity of sex expression provides the base for the future breeding programs. Several gaps in traditional plant breeding methods like development of a stable CMS system, resistance to Botrytis gray mold rot, and the complexity of sex reversion can only be addressed through collaborative and multidisciplinary approach.

References

Alexandrov OS, Karlov GI (2016) Molecular cytogenetic analysis and genomic organization of major DNA repeats in castor bean (*Ricinus communis* L.). Mol Genet Genom 291(2):775–787

Ananthasayana K, Reddy MV (1981) Correlations and path analysis of yield components in inbreds and hybrids of castor (*Ricinus communis* L.). Andhra Agric J 28:112–115

Anastasi U, Sortino O, Cosentino SL, Patanè C (2015) Seed yield and oil quality of perennial castor bean in a Mediterranean environment. Intl J Plant Prod 9(1):99–116

Anjani K (1997) Inheritance of spininess of capsules in castor (*Ricinus communis* L.). Crop Improv 24 (1):113–114

Anjani K, Pallavi M, Babu SNS (2007) Uniparental inheritance of purple leaf and the associated resistance to leafminer in castor bean. Plant Breeding. https://doi.org/10.1111/j.1439-0523.2007.01395.x

Ankineedu G, Rao NGP (1973) Development of pistillate castor. Indian J Genet 33:416–422

Anonymous (2006) AICRP on Castor, IIOR, Research achievements in castor, ICAR-IIOR, Hyderabad, India, 111 p

Anonymous (2007) Department of Science and Technology, J. V. Seriña St, Carmen, Cagayan de Oro

Anonymous (2017) ICAR-IIOR, Annual Report, 2016–17, ICAR-IIOR, Hyderabad, India

Atsmon D (1989) Castor. In: Robbelen G, Downey RK, Ashri A (eds) Oil crops of the world. McGraw Hill, New York, pp 348–447

Auld DL, Rolfe RD, McKeon TA (2001) Development of castor with reduced toxicity. J New Seeds 3:61–69

Auld DL, Zanotto MD, Mckeon TA, Morris JB (2009) Castor. In: Vollman J, Rajcan I (eds) Oil crops breeding. Handbook of plant breeding. Springer, New York, pp 316–332

Barad YM, Pathak AR, Patel BN (2009) Heterosis for yield attributing traits in castor (*Ricinus communis* L.). J Oilseeds Res 26(2):102–103

Bhapkar DG, D'Cruz R (1967) Inheritance of number of main leaf veins in castor. Indian J Genet 27(1):151

Bhardwaj HL, Mohamed AI, Webber CL III, Lovell GR (1996) Evaluation of castor germplasm for agronomic and oil characteristics. In: Janick J (ed) Progress in new crops. ASHS Press, Alexandria, VA, pp 342–346

Bhatt D, Reddy TP (1986) Genetics of seed yield and its components in castor (*Ricinus communis* L.). Genet Agrar 40(2):193–204

Bhatt D, Reddy TP (1981) Correlations and path analysis in castor (*Ricinus communis* L.). Can J Genet Cytol 23:523–531

Bhatt D, Reddy TP (1983) Combining ability for flowering capsule number and yield in castor. Indian J Agric Sci 53:873–875

Brigham RD (1967) Natural outcrossing in dwarf-internode castor *Ricinus communis* L. Crop Sci 7:353–354

Brigham RD (1993) Castor: return of an old crop. In: Janick J, Simon JE (eds) Progress in new crops. Wiley, New York, pp 380–383

Brown SW (1949) The structure and meiotic behavior of the differentiated chromosomes of tomato. Genetics 34 (4):437

Chakrabarti MH, Ahmad R (2008) Trans-esterification studies on castor oil as a first step towards its use in biodiesel production. Pak J Bot 40(3):1153–1157

Chakrabarty SK (1997) Combining ability and heterosis studies in castor (*Ricinus communis* L.). J Oilseeds Res 14:182–188

Chakrabarty SK, Banu SS (1999) Genetic diversity in pistillate lines of castor. Indian J Agric Sci 69 (10):729–730

Chakrabarty SK, Lavanya C, Mukta N (2006) Draft National guidelines for the conduct of tests for distinctness, uniformity and stability castor (*Ricinus communis* L.) 28 p

Chakrabarty SK, Lavanya C, Mukta N (2009) Castor. In: Dinesh Kumar, Mandal AB (eds) National guidelines for the conduct of tests for distinctness, uniformity and

stability. Directorate of Seed Research, Kushmaur, MAU, UP, 440 p
Chandramohan Y (2002) Genetic analysis of some qualitative and quantitative traits in castor (ricinus communis L). Ph d thesis ANGRAU, Hyderabad, India
Chandrasekharan SN, Sundaraj DDA (1946) Note on the infloresence of *Ricinus communis* L. J Indian Bot Soc 25:103
Chauhan SS, Saxena BK, Kinoshita T (1987) Effect of daminozide (B9) on sex-expression and seed setting in castor bean, *Ricinus communis* L. Jpn J Breed 37 (3):262–266
Chauhan SV, Singh KP, Kinoshita T (1990) Gamma-ray induced pollen sterility in castor (*Ricinus communis*). J Fac Agri-Hokkaido Univ 64(3):229–234
Chauhan SVS, Singh KP, Saxena BK (1992) Gamma-ray induced female mutation in castor. Indian J Genet 52:26–28
Claassen CE, Hoffman A (1950) The inheritance of the pistillate character in castor and its possible utilization in the production of commercial hybrid seed. Agron J 42:79–82
Conceição MM, Fernandes VJ Jr, Bezerra AF, Silva MC, Santos IM, Silva FC, Souza AG (2007) Dynamic kinetic calculation of castor oil biodiesel. J Therm Anal Calorim 87(3):865–869
Costa MN, Pereira WE, Bruno RLA, Freire EC, Nóbrega MBM, Milani M, Oliveira AP (2006) Genetic divergence on castor bean accesses and cultivars through multivariate analysis. Pesqi Agropecu Bras 41:1617–1622
da Silva Ramos LC, Tango JS, Savi A, Leal NR (1984) Variability for oil and fatty acid composition in castorbean varieties. J Am Oil Chem Soc 61 (12):1841–1843
Damodaram T, Hegde DM (2010) Oilseeds situation, a statistical compendium. Directorate of Oilseeds Research, Hyderabad, India
Dangaria GJ, Dobaria KL, Fatteh UG, Patel VJ (1987) Heterosis and combining ability analysis in castor. J Oilseeds Res 4:46–53
Desai AG, Dange SRS, Pathak HC (2001) Genetics of resistance to wilt in castor caused by *Fusarium oxysporum* f. sp. *ricini*. Nanda and Prasad. J Mycol Plant Pathol 31(3):322–326
Dhapke SK, Khorgade PW, Narkhade MN (1992) Estimates of genetic variability in castor (*Ricinus communis*). Agric Sci Digest (Karnal) 12:141–143
Dobariya KL, Dangaria GJ, Patel VJ (1992) Combining ability analysis in a castor diallel. J Maharash Agric Univ 17:235–238
Dobariya KL, Patel ID, Patel PS, Patel VJ (1989) Combining ability and genetic architecture of oil content in castor (*Ricinus communis* L.). J Oilseeds Res 6(1):92–96
Dorairaj S, Kandawami M, Palaniswamy S, Varisaimuhammed S (1973a) Correlation studies in *Ricinus communis* L. I with inbreds and hybrids. Madras Agric J 60:1481–1485
Dorairaj SM, Kandasami M, Palaniswamy S, Varisai S (1973b) Correlation studies in *Ricinus communis* L. within inbreds and hybrids. Madras Agric J 60:1486–1490
Fatteh UG, Dangaria CJ, Doariya KL, Patel VJ (1987) Combining ability for seed characters in castor. J Oilseed Res 4:54–58
Fatteh UG, Dangaria CJ, Dobariya KL, Patel VJ (1988) Combining ability by line × tester analysis in castor. Indian J Agric Sci 58:7–10
Filho SA (1999) Improvement of castor. In: The Borém (ed) Improvement of cultivated species. The Borem Federal University of Vicosa, Vicosa, Brazil
Gedil M, Kolade F, Raji A, Ingelbrecht I, Dixon A (2009) Development of molecular genomic tools for verification of intergeneric hybrids between castor bean (*Ricinus communis L.)* and cassava (*Manihot esculenta* Crantz). J Food Agric Environ 7(2):534–539
George WL (1966) Genetics of interspersed staminate flowers in *Ricinus communis* L. Diss Abstr 38:66
George WL, Shifriss (1967) Interspersed sexuality in castor. Genetics 57:347–356
Giriraj K, Mensinkai SW, Sindagi SS (1973) Correlation and genetic studies in diallel crosses of castor (*Ricinus communis* L.). Mysore J Agric Sci 7:385–388
Giriraj K, Mensinkai SW, Sindagi SS (1974) Components of genetic variation for yield and its attributes in 6 × 6 diallel crosses of castor (*Ricinus communis* L.). Mysore J Agric Sci 44:132
Golakia PR, Kavani RH, Madaria RB, Monpara BA, Savalia RL (2005) Reaction of newly developed pistillate and inbred lines and their hybrids of castor to *Fusarium oxysporum*, f.sp. *ricini.* Nanda and Prasad. J Oilseeds Res 22:422–423
Gopala Rao P (1969) Sex expression in *Ricinus communis* L. Sci Cult 40:326–327
Gopani DD, Kabaria MM, Patel RH (1969) Study of sex reversion in castor. Indian J Agric Sci 40:255–258
Gopani DD, Kabaria MM, Patel RH (1970) Effect of spacing, fertilizer and plants per hill on sex variation in castor *Ricinus communis* L. Indian J Agric Sci 40 (1):17–23
Gourishankar V, Rao PVR, Reddy AVV (2010) Inheritance of certain morphological characters and Fusarium wilt resistance in castor, *Ricinus communis* L. SABRAO J Breed Genet 42(2):57–64
Harland SC (1920) Inheritance in *Ricinus communis* L. Part I. J Genet 10(3):207–218
Harland SC (1922) Inheritance in *Ricinus communis* L. Part II. J Genet 12:251–253
Hawkins R (1989) Bring back the US castor crop. J Am Chem Soc 66:1567–1568
Hegde DM, Lavanya C (2012) Chapter 19. Castor. In: Bharadwaj DN (ed) Breeding of field crops. AGROBIOS INDIA. ISBN978-81-7754-474-9. pp 471–512
Hilterbrandt VM (1935) The plant resources of the world as initial material in plant breeding. Castor Plant 55–70 Lenin. Acad Agric Sci Inst Plant Indust, Moscow and Leningrad, No 6

ICOA (1992) The chemistry of castor oil and its derivatives and their applications. Tech Bull No. 2, International Castor Oil Association

Jaiswal VS, Kumar A (1983) Peroxidase activity and its isozymes in relation to flower sex expression in *Ricinus communis* L. Short communication. Curr Sci 52:368–369

Jakob KM (1956) The pachytene chromosomes of the castor oil plant. Cytologia 21:76–80

Jelenkovic G, Harrington E (1973) Chromosome complement of *Ricinus communis* at pachytene and early diplotene. J Hered 64:137–142

Joshi WV (1926) Some variation in the inflorescence of castor plant. Poona Agric College Mag 18(1):26–29

Kandasamy M, Palanisamy S, Dorairaj MS (1983) Combining ability in *Ricinus communis* L. Madras Agric J 70:755–757

Kaul S L, Prasad MVR (1983) Genetic basis of yield heterosis in castor. In: Pre-congress scientific meeting on genetics and improvement of heterotic systems, Coimbatore, India, vol 54, pp 15–16

Konwar PC (1960) Observations of hermaphroditism in the flowers of *Ricinus communis* L. Sci Cult 26:83

Kulkarni LG, Ankineedu G (1966) Isolation of pistillate lines in castor for exploitation of hybrid vigour. Iindain J Genet 26:363–365

Kulkarni LG, Ramanamurthy GV (1977) Castor. ICAR, New Delhi, p 105

Lakshmamma P, Lakshmi P, Anjani K (2004) Screening of castor germplasm accessions for drought tolerance. Indian J Plant Physiol 9:447–450

Lakshmamma P, Lakshmi P, Mohan YC, Lavanya C (2005) Genetic variability and character association in castor (*Ricinus communis* L.). Natl J Plant Improv 7 (2):122–126

Lakshmamma P, Prayaga L (2006a) Identifying the sources of tolerance for drought in castor, *Ricinus communis* L. J Oilseeds Res 23(2):348

Lakshmamma P, Prayaga L (2006b) Screening castor, *Ricinus communis* L. germplasm lines for thermal tolerance. J Oilseeds Res 23(2):353

Lakshmamma P, Prayaga L (2010) Variability for water use efficiency traits and drought tolerance in castor (*Ricinus communis* L.) germplasm lines. J Oilseeds Res (in press)

Lakshmamma P, Prayaga L, Lavanya C (2002) Physiological manipulation of sex expression in a pistillate line of castor. J Oilseeds Res 19(1):104–106

Lakshminarayana G, Paulose MM, Kumari BN (1984) Characteristics and composition of newer varieties of Indian castor seed and oil. J Am Oil Chem Soc 61 (12):1871–1872

Laureti D (1981) Comparative trial of *Ricinus* cultivars. Sementi Elette 27:23–28

Lavanya C (2002) Sensitivity and variations of sex expression in response to environmental changes in castor. Indian J Genet 62(3):232–237

Lavanya C, Anjani K, Rao G (2006) N.V.P.R. 2006. Crop improvement. In: Hegde DM (ed) Research achievements in castor. All India coordinated research project on castor. Directorate of Oilseeds Research, ICAR, pp 8–37

Lavanya C, Chakrabarthy SK, Ramachandram M, Rao CH, Raoof MA (2003) Development of wilt resistant pistillate lines in castor through mutation breeding. J Oilseeds Res 20(1):48–50

Lavanya C, Dudhe MY, Zaveri PP, Meena HP (2010) Exploration of cytoplasmic male sterility sources in castor—a concept note. J Oilseeds Res 27(Spl Iss):164–166

Lavanya C, Gopinath V (2008) Study on inheritance of morphological characters and sex expression in pistillate lines of castor (*Ricinus communis* L.). Indian J Genet 68(3):275–282

Lavanya C, Lakshmamma P (2008) Selection of high-yielding lines in castor (*Ricinus communis*). Indian J Agric Sci 78(4):318–322

Lavanya C, Mohan Y (2003) Combining ability and heterosis for seed yield and yield components in castor. J Oilseeds Res 20(2):220–224

Lavanya C, Murthy GSS, Lakshminarayana M (2008) Use of gamma rays for development of leaf hopper resistant pistillate lines in castor (*Ricinus communis* L.). In: Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, Plant Breeding Section, Vienna (Austria) (2008). International symposium on Induced Mutations in Plants (ISIM) Book of Abstracts (IAEA-CN–167). International Atomic Energy Agency (IAEA), Viena, Austria, 168 p

Lavanya C, Ramanarao PV, Gopinath V (2006b) Studies on combining ability and heterosis in castor hybrids. J Oilseeds Res 23(2):174–177

Lavanya C, Rao HC, Chakrabarthy SK, Ramachandram M, Raoof MA, Prasad MVR (2000) Use of mutation breeding in hybrid castor development. In: Proceedings of DAE-BRNS symposium on use of nuclear and molecular techniques in crop improvement. BARC, Mumbai, India, pp 194–202

Lavanya C, Raoof MA, Santa LP (2011) Genetics of resistance to *Fusarium* wilt in castor caused by *Fusarium oxysporum* f. sp. *ricini* Nanda and Prasad. Indian Phytopathol 64(2):151–153

Lavanya C, Anjani K, Ganga Rao NVPR, Lakshmamma P (2006) Research achievements A. Crop improvement. In: Hegde DM (ed) AICRP on castor. DOR 2006 research achievements in castor. All India coordinated research project on castor, Directorate of Oilseeds Research, Hyderabad, India, pp 8–37

Lavanya C, Reddy AVV, Gangarao NVPR, Dineshkumar V (2006) Research achievements D. Seed production technology. In: Hegde DM (ed) AICRP on castor. DOR 2006 research achievements in castor. All India coordinated research project on castor, Directorate of Oilseeds Research, Hyderabad, India, pp 81–90

Lavanya C, Solanki SS (2010) Crop improvement of castor. The challenges ahead. In: Hegde DM (ed) Research and development in castor. Present Status and Future Strategies. Indian Society of Oilseeds Research Hyderabad, pp 36–55

Lavanya C, Ushakiran B, Manjunatha T, Santha MSLP, Reddy AVV (2018) Development of diverse parental lines in castor through single seed descent method. In: Dey R, Ajay BC, Mahatma MK, Thirumalaisamy PP, Gangadhara K, et al. (eds) In: Extended summaries: national conference of oilseeds (NCOS2018), 7–9 Apr 2018, ICAR-DGR, Junagadh, Gujarat, India, pp 39–40

Lavanya C, Varaprasad KS (2012) Castor hybrids in India: a success story. Seed Times 5(4):111–117

Lavanya C, Duraimurugan P, Manjunatha T, Senthilvel S, Ramy KT, Santha MSLP, Lakshmamma P, Reddy AVV (2018) Development of leafhopper and wilt resistant pistillate line in castor through mutation breeding. In: International symposium on plant mutation breeding and biotechnology (ISPMBB), International Atomic Energy Agency (IAEA), Vienna, Austria, 27–31 August, 2018 (in press)

Loui IH, Dorai Raj MS, Selvaraj J (1986) Induced female sterility in castor. Madras Agric J 73:643–644

Madariya RB, Vadderia MA, Mehta DR, Kavani RH (2008) Combining ability analysis over environments for seed yield and its components in castor (*Ricinus communis* L.). Crop Improv 35(2):163–166

Manivel P, Hussain HS (1997) Genotype × environment interaction for combining ability estimates in castor (*Ricinus communis* L.). J Oilseeds Res 14:31–35

Manivel P, Hussain JHS, Raveendran TS (1998) Combining ability for earliness traits over environments in castor. Madras Agric J 85:157–159

Mehta DR, Vashi PS, Kukadia MV (1991) Combining ability for earliness and its related traits in castor (*Ricinus communis* L.) Gujarat Agrl Univ J (17):16–21

Mehta DR, Vashi PS (1998) Correlation and path analysis of seed yield and its components in castor. Indian J Agric Res 32(3):160–164

Mohan YC, Reddy AV, Rao TN (2006) Inheritance of bloom nature in castor, *Ricinus communis* L. J Oilseeds Res 23(2):184

Moshkin VA (1986) Cytology and genetics of qualitative characteristics. In: Moshkin VA (ed) Castor. Amerind Publication, Delhi, India, pp 125–132

Mukta N, Reddy AP, Lavanya C, Dudhe MY (2012) Variability for DUS characteristics in released varieties of safflower (*Carthamus tinctorius* L.) in India. J Oilseeds Res 29:133–135

Murthy GK, Reddy AVV, Lakshmamma P, Rao NT (2003) Manipulation of sex expression in castor (*Ricinus communis* L.) by chemical treatments. J Oilseeds Res 20(2):225–228

Muthiah A, Manivannan M, Natarajan C, Palanisamy GA (1982) Correlation and path analysis in *Ricinus communis*. Madras Agric J 69:561–563

Narain A (1952) Bloom characters in castor oil plant (*Ricinus communis* L). Curr Sci 21(6):166–167

Narain A (1961) Inheritance of bloom in castor. Indian J Genet 21:136–141

Narian A (1951) Mutants in castor oil plant. Sci Cult 16:484–485

Narian A (1953) Artificial production of tetraploids in *Ricinus communis* L. Curr Sci 22(9):268–269

Narkhede BN, Jadhav MS, Deokar AB (1985) Combining ability analysis in castor. J Maharash Agric Univ 10:102–103

Neeraja P, Lavanya C, Ganesh M (2010) Quality hybrid seed production in castor by manipulation of pistillate character. J Oilseeds Res 27(Spl Iss):130–132

Oliveira IJ, Zanotto MD (2008) Efficiency of recurrent selection for reduction of the stature of plants in castor (*Ricinus communis* L.). Ciência Agrotecnol 32:1107–1112

Omkarappa T, Hanumanthappa DC, Kumar A (2010) Correlation on path analysis of yield attributing traits in hybrids of castor (*Ricinus communis* L.) J Oilseeds Res 27(Spl Iss):72–74

Padhar PR, Chovatia VP, Barad VG, Patel MB, Dobariya KL (2010) Combining ability by line × tester analysis in castor (*Ricinus communis* L.). J Oilseeds Res 27(Spl Iss): 44–46

Paris HS (1981) Pachytene chromosome morphology with reference to sex instability in *Ricinus communis* L. Euphytica 30:193–196

Paris HS, Shifriss O, Jelenkovic G (1978) Idiogram of *Ricinus communis* L. J Hered 69(3):191–196

Parkey W (1957) Cytoplasmic Influence in the production of the pistillate sex expression in castor beans 1. Agron J 49(8):427–428

Patel DK, Patel CJ, Patel PJ (2010) Study of combining ability for yield and its components in castor (*Ricinus communis* L.) through line × tester analysis. J Oilseeds Res 27:51–53

Patel ID, Dangaria CJ, Dholaria SJ, Patel VJ (1984) Phenotypic stability of seed yield in eight hybrids of castor. Indian J Agric Sci 54(8):629–632

Patel ID, Dangaria CJ, Patel VJ (1986a) Induction of male sex in pistillate lines of castor. Indian J Agric Sci 56:556–558

Patel PS, Jaimini SN (1991) Inter-relationship and path analysis of certain quantitative characters in castor (*Ricinus communis* L.). J Oilseeds Res 8(1): 105–109

Patel RH, Joshi SN (1972) Note on the influence of sex characters on seed production of hybrid castor. Indian J Agric Sci 42:647–648

Patel VJ, Dangaria GJ, Fatteh UG, Patel PS (1986b) Line × tester analysis of combining ability in castor. J Oilseeds Res 3:178–183

Patel PB, Pathak HC (2011) Genetics of resistance to wilt in castor caused by *Fusarium oxysporum* f.sp. *ricini* Nanda and Prasad. Agric Sci Dig 31(1):30–34

Pathak GN, Asthana AN (1962) Distribution of red pigments in castor. Indian Oilseeds J 5(2):84–88

Pathak GN, Dikshit NN, Vaish CP (1965) A note on the occurrence of anomalous capsules in castor. Indian Oilseeds J 9:157

Pathak HC, Dangaria CJ (1987) Combining ability for yield and its components in castor. Indian J Agric Sci 57:13–16

Pathak HC, Dangaria CJ, Parmar KS (1986) Heterosis and genetic architecture of oil content in castor (*Ricinus communis* L.). Madras Agric J 73(6):328–333

Pathak HC, Dixit SK, Patel PG (1988) Gene effects and heterosis in castor (*Ricinus communis* Linn). Indian J Agric Sci 58:476–478

Patwardhan GW (1931) A preliminary note on inheritance of castor. J Ind Bot Soc 10:100–109

Peat JE (1926) Genetic studies in *Riciuns communis* L J. Genet 19:373–89

Pinkerton SD, Rolfe R, Auld DL, Ghetie V, Lauterbach BF (1999) Selection of castor for divergent concentrations of ricin and *Ricinus communis* agglutinin. Crop Sci 39(2):353–357

Podkuichenko GV (1989) Donors of resistance to *Fusarium* in castor oil plant. Instituta Maslichnykh Kullar 1:24–26

Prayaga P, Lakshmamma P (2009) New technique for root studies. DOR Newsl 14(4):1–2

Pushpavathi B, Sarwar HAK, Raoof MA, Babu RR (1998) Management of wilt disease in castor. Indian J Plant Prot 26(2):177–180

Raghavaiah CV, Lavanya C, Kumaran S, Royal TJJ (2006) Screening castor (*Ricinus communis*) genotypes for salinity tolerance in terms of germination, growth and plant ion composition. Indian J Agric Sci 76(3):196–199

Raghavaiah CV, Muralidharudu Y, Ammaji P, Royal TJ, Padmaiah M, Lavanya C, Lakshmamma P (2003) Castor production in relation to certain agronomic measures under sodic vertisol conditions. J Oilseeds Res 20(2):287–291

Raghavaiah CV, Muralidharudu Y, Joseph Jeevan Royal T, Ammaji P, Lavanya C, Lakshmamma P (2002) Influence of salinity stress on germination and early growth of castor (*Ricinus communis*) genotypes. Indian J Agric Sci 72(10):601–603

Raju SCH (1981) Studies on character associations in hybrid of castor (*Ricinus communis* L.). M.Sc. (Ag) thesis, ANGRAU, Rajendranagar, Hyderabad, India

Rakshit S (2008) Heterotic gene pool development in maize. Lecture given in the Winter School on Innovative approaches for increasing productivity in oilseed crops—a crop improvement perspective, 4–24 Nov, DOR, Hyderabad, India

Ram HM, Sett R (1980) Induction of male flowers in a pistillate line of *Ricinus communis* L. by silver and cobalt ions. Planta 149(4):413–415

Ramesh M, Lavanya C, Rao NVPRG, Rao BMV (2010) Studies on genotypic correlations and path coefficient analysis of seed yield and yield components in castor genotypes. J Oilseeds Res 27(Spl Iss):128–130

Ramesh M, Lavanya C, Rao MVB, Meena HP (2012) Genetic variability, heritability and genetic advance for yield and yield components in castor (*Ricinus communis* L.) genotypes. J Oilseeds Res 29(1):101–103

Ramesh T, Reddy DVV, Reddy TV, Prasad MMKD (2000) Effect of ethrel on sex expression and yield of castor (*Ricinus communis* L.). J Oilseeds Res 17 (2):279–281

Ramesh T, Venkateswarlu O, Prasad MMKD (2001) Path analysis in castor. Agrl Sci Digest 21(1):59–60

Ramu R, Sreedhar N, Lavanya C, Ramesh T (2012) Combining ability studies in castor. J Oilseeds Res 19 (2):229–230

Rao BS, Rayan T (1954) A note on the abnormality in the fruits of castor. Madras J Agric 41:228

Rao CH, Raoof MA, Lavanya C (2005) Study on segregation patterns and linkages between morphological characters and wilt resistance in castor (*Ricinus communis* L). J Oilseeds Res 22(1):114–118

Rao CH, Lavanya C, Chakrabathy SK (2000) Diversification of pistillate base in castor In extended summaries 2000 National Seminar on Oilseeds and Oils-Research and development needs in the millennium February 2–4 2000. Indian Society of Oilseeds Research, DOR, Hyderabad, pp 88–90

Ratnakar B (1982) Variability and character association in castor (*Ricinus communis* L.). Andhra Agric J 29:194–197

Ratnakumari P (1996) Stability analysis in castor (*Ricinus communis* L.) M.Sc. (Ag) Thesis, Acharya NG Ranga Agricultural University, Hyderabad, India

Reddy PRR, Sankar GRM, Vanaja M, Rao CH, Venkateswarlu S, Eastin JD (1999) Heritability and genetic advance in castor (*Ricinus communis*) over varied soils and dates of sowing. Indian J Agric Sci 69 (5):371–375

Richharia RH (1937) A note on cytogenetics of *Ricinus communis* L. Indian J Agric Sci 7(5):707–711

Rojas-Barros P, de Haro A, Munoz J, Fernandez-Martinez JM (2004) Isolation of a natural mutant in castor with high oleic/low ricinoleic acid content in the oil. Crop Sci 44(1):76–80

Saiyed MP, Shukla PT, Dhameliya H (1997) Heterosis for yield and yield components in castor (*Ricinus communis* L.) over environments. Gujarat Agric Univ Res J 23:22–27

Santha MSLP, Bharathi E, Lavanya C, Prabakharan AJ (2016) Parental lines and advanced breeding material of castor resistant to wilt disease. Indian Phytopathol 69(4s):721–723

Santha MSLP, Raoof MA, Giribabu P, Anjani K, Lavanya C, Varaprasad KS (2014) Castor wilt. A monograph. ICAR-Directorate of Oilseeds Research, Hyderabad, India, 34 p

Satyabalan K, Ponnaiya BW, Raman VS (1965) Studies on hybrid vigour in castor. Indian Oilseeds J 9(3):283–289

Savithri V (1963) A case of phyllody in the inflorescence of *Ricinus communis* Linn. Madras Agric J 50:480

Savithri V, Aiyadurai SG (1961) Occurrence of synocotylous seedlings in *Ricinus communis* L. Madras Agric J 48:148

Seshadri CR, Muhammad V (1951) Study of inheritance of certain characters. In: Castor-proc first scientific workers' conference, Agricultural College and Research Institute, Coimbatore, Madras, India

Seshadri CR, Seshu KA (1956) Preliminary observation on Jassid injury in castor. Madras Agric J 43:197–199

Severino LS, Auld DL, Baldanzi M, Cândido MJ, Chen G, Crosby W, Tan D, He X, Lakshmamma P, Lavanya C, Machado OL (2012) A review on the challenges for increased production of castor. Agron J 104(4):853–880

Shifriss O (1960) Conventional and unconventional systems controlling sex variations in *Ricinus*. J Genet 57:573–578

Shifriss O (1964) Growth and sexuality of *Ricinus communis* L. in a constant environment. Am Nat 98:187–189

Shinde VS, Pawar KR, Chavan BN (1985) Correlation and regression coefficient studies in castor. J Maharash Agric Univ 10:354–355

Sindagi SS (1964) Castor hybrids I. Manifestation of hybrid vigour in F1s during early seedling stages. Indian Oilseeds J 8 (2):159–161

Singh A, Srivastava AN (1982) Note on line × tester analysis in castor. Indian J Agric Sci 52:610–612

Singh H, Yadava TP (1981) Genetic analysis of days to flowering, maturity and yield in castor. J Res Haryana Agric Univ (Hisar) 11:54–59

Singh SB, Singh HG, Singh RR (1981) Path coefficient analysis in castor. Madras Agric J 68:688–689

Solanki SS, Joshi P (2000) Combining ability analysis over environments of diverse pistillate and male parents for seed yield and other traits in castor (*Riciuns communis* L.). Indian J Genet 60:201–212

Solanki SS, Joshi P (2001) Inheritance study of morphological traits in castor (*Riciuns communis* L.). Indian J Genet 61:136–139

Sviridov AA (1988) Results of improving the parent forms of the castor hybrid Kranoda rskit 3 for resistance to *Fusarium*. Institute Maslichnykh Kultur. 2:16–19

Varaprasad KS, Lavanya C (2015) Innovative approaches to break yield barrier in oilseeds and improved strategies in quality seed supply. In: Sultana R, Durga KK, Lakshmi YS, Sreedhar M, Rao LVS Keshavulu K (eds) In: Compendium of lead papers and abstracts, 8th National Seed Congress, 2015. pp 35–41

Varisai M (1970) Hybrid vigour in *Ricinus communis* L. II. General combining ability of certain exotic and indigenous inbreds with TMV. Madras Agric J 57:578–584

Vasconcelos S, De Souza A, Gusmao CLS, Milani M, Ana MBI, Ana CBV (2010) Heterochromatin and rDNA 5S and 45S sites as reliable cytogenetic markers for castor bean (*Ricinus communis*, Euphorbiaceae). Micron 41:746–753

Varisai S, Srinivasalu N, Dorairaj MS, Navakodi K, Ramaswamy NM, Sivaram MR (1969) Hybrid vigour in *Ricinus communis* LI General combining ability between certain Egyptian inbreds and TMV 1. Madras Agric J 40:244–245

Wang ML, Morris JB, Pinnow DL, Davis J, Raymer P, Pederson GA (2010) A survey of the castor oil content, seed weight and seed-coat colour on the United States Department of Agriculture germplasm collection. Plant Genet Resour 8(3):229–231

Weibe RO, Woodworth CM (1946) Use of natural crossing plot in making castor bean hybrids. J Am Soc Agric 38(6):563

Weiss EA (1971) Castor, sesame and safflower NeDA. New York

Weiss EA (2000) Oilseed crops, 2nd edn. Blackwell Science, Oxford, p 364

White OE (1918a) Breeding new castor beans. J Hered 9:195–200

White OE (1918b) Inheritance studies on castor beans. Brooklyn Bot Garden Mem 1:513–521

Yadava YP, Yadav IS, Ravi P, Singh R, Yadava JS (2004) Genotypic association and path coefficient analysis in castor (*Ricinus communis* L.). J Oilseeds Res 21(2):346–348

Zaveri PP, Lavanya C, Hegde DM (2010) Public private partnership for increasing productivity in castor. In: Hegde DM (Ed) Research and development in castor: present status and future strategies. Indian Society of Oilseeds Research, Hyderabad, pp 210–224

Zhong XB, Fransz PF, Wennekes-van Eden J, Zabel P, Van Kammen AB (1996) High-resolution mapping on pachytene chromosomes and extended DNA fibres by fluorescencein-situ hybridisation. Plant Mol Biol Rep 14(3):232–242

Zhou GL, Gu MX (1990) Selection of a marked female line of castor bean (*Ricinus communis*) and a study of its inheritance Oil Crops of China No. 3 pp. 23–26. ISSN 1000-6303. Record Number 19911622133

Zimmerman LH, Parkey W (1954) Pistillate F_1 castor beans: their possible significance in producing commercial hybrid seed 1. Agron J 46(6):287

Recurrent Selection for Improved Oil Content in Castor Bean

4

Grace Q. Chen, Kumiko Johnson and Eva Morale

Abstract

Castor (*Ricinus communis*) seed oil is currently the only commercial source of hydroxy fatty acid (12-hydroxyoleic acid, 18:1OH) which is used in industrial products such as lubricants, coatings, plastics, and cosmetics. Increasing seed oil content has immense benefits as it is one of the major factors contributing to oil yield. This chapter describes an effective procedure to increase oil content through recurrent selection based on screening individual castor seeds using nuclear magnetic resonance technology.

4.1 Introduction

Castor bean (*Ricinus communis*) is an important oilseed crop in the Euphorbiaceae family. Castor oil contains 90% an uncommon hydroxylated fatty acid, ricinoleic acid (or 12-hydroxyoleic) (Wang et al. 2011). The hydroxy group imparts unique chemical and physical properties that make castor oil a vital raw material for manufacturing numerous industrial products, such as high lubricity lubricants and greases, a variety of polymers traditionally made by petroleum-based products for various coatings, elastomers and plastics, and an array of castor oil derivatives utilized in cosmetic industry (Caupin 1997; Sharma and Kundu 2006; Johnson 2007; Mutlu and Meier 2010). Besides, several favorable features, including high oil content (37–60% of seed dry weight) (Wang et al. 2010), high oil yield (1250–2500 L ha^{-1}) (Severino et al. 2012), and being a nonfood crop, make castor bean an attractive industrial oilseed.

Improvement of castor bean cultivars by increasing oil content would be of great benefit as it contributes to oil yield and makes the production more cost-competitive. Sprague et al. (1952) demonstrated that the oil content can be increased through recurrent selection. The application of nuclear magnetic resonance spectroscopy, hereinafter designated "NMR," to nondestructively achieve analysis of oil in living seeds (Conway and Earle 1963; Rubel 1994), or even in a single seed (Bauman et al. 1963), allows quick selection of seeds and the selected seeds can be directly planted after measurement. Single-seed recurrent selection has been demonstrated to be an effective method to increase seed oil content in a number of species such as maize (Lucas et al. 2013), soybean (Feng et al. 2004), and oat (Frey and Holland 1999).

G. Q. Chen (✉) · K. Johnson · E. Morale
Agricultural Research Service, US Department of Agriculture, Western Regional Research Center, 800 Buchanan Street, 94710 Albany, CA, USA
e-mail: grace.chen@ars.usda.gov

C. Kole and P. Rabinowicz (eds.), *The Castor Bean Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97280-0_4

4.1.1 Using Nuclear Magnetic Resonance Technology for Recurrent Selection

A nuclear magnetic resonance (NMR) analyzer can be used to measure the percentage of oil in a living castor bean seed (Chen et al. 2016). Each single castor bean seed is weighed and placed in an NMR tube and then measured against a calibration curve to determine oil content. Recurrent selection is performed by the following basic procedures: (1) individual seeds with high oil content are selected from a base population by screening the base population using NMR; (2) selected seeds are planted and a new generation of seeds is produced through open-pollination; (3) the new generation of seeds forms a new population for the next cycle of screening and selection.

4.1.2 A Test Case of Recurrent Selection

We performed two cycles of recurrent selection using an NMR analyzer in a greenhouse with controlled temperatures ranging from 18 to 28 °C (night/day) under a 15/9 h (day/night) photoperiod. A base population consisted of bulk Impala cultivar seeds (Chen et al. 2016). Individual seeds were germinated in 4" pots containing peat-vermiculite growth mixture and seedlings with four true leaves were transferred to five-gallon pots supplied with complete fertilizer throughout the remaining time of plant growth.

The schematic selection procedures were as follows:

Cycle 0 seeds or base population:	233 seeds screened for oil content using NMR
	↓
	Top 30 seeds selected, planted, and produced
	Cycle 1 seeds (or population) in a greenhouse
	↓
Cycle 1 population:	Total 770 seeds screened for oil content using NMR
	↓
	Top 30 seeds selected, planted, and produced
	Cycle 2 seeds (or population) in a greenhouse
	↓
Cycle 2 population:	Total 1859 seeds measured for oil content

The results of the recurrent selection are summarized in Table 4.1 and Fig. 4.1. The oil content in the base population ranged from 40.13 to 57.49% with an average of 50.33%. The top 30 seeds with an oil content of 53% or higher were planted in a greenhouse and 770 mature seeds (or Cycle 1) were collected and analyzed for oil content. The oil content in Cycle 1 ranged from 44.34 to 58.65% with an average of 53.87%. The shift in the mean oil content of Cycle 1 showed a 3.54% increase, with a 95% lower confidence limit of 3.28%. To further improve seed oil content, the top 30 seeds of Cycle 1 with an oil content of 56% or greater were planted in the greenhouse and 1859 s cycle seeds (Cycle 2) were collected. The oil content in Cycle 2 ranged from 45.26 to 59.16% with an average of 54.47%. Cycle 2 exhibited an increase in mean oil content of 4.14% compared with that of Cycle 0. From Cycle 1 to Cycle 2, a 0.6% increase in mean oil content was observed. In addition to the increase of mean oil content, we also observed a reduction of natural variability in oil content after the selection. The original base population had a standard deviation of 2.98%, which was reduced to 1.93% in Cycle 1 and 1.62% in Cycle 2. Besides, the range of seed oil content (minimum-to-maximum) was also reduced from 17.36% in Cycle 0 to 14.31% in Cycle 1 and 13.9% in Cycle 2. Therefore, the average seed oil content was increased because the frequency of seeds with low oil content was reduced and the frequency of seeds with high oil content was increased after selection (Fig. 4.1). We did not observe any adverse effect on plant

Table 4.1 Cycle means, standard deviation (SD), minimum (min), maximum (max)

Population	Number of seeds	Oil content (%)	
		Mean ± SD	Min–max (%)
Cycle 0	233	50.33 ± 2.98	40.13–57.49
Cycle 1	770	53.87 ± 1.93	44.34–58.65
Cycle 2	1859	54.47 ± 1.62	45.26–59.16

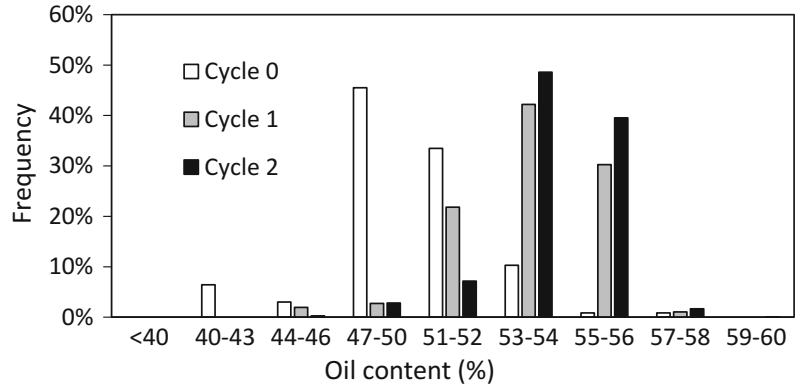

Fig. 4.1 Comparison of frequency distributions of oil content in cycle 0, cycle 1, and cycle 2 of individual castor seeds

growth and morphology during our selection. In the greenhouse, castor pollen can be transferred from the anther of a male flower to the stigma of a female on the same plant (self-pollination), or to the stigma on a different plant (cross-pollination). Such mixed pollination system apparently avoided inbreeding depression.

Based on the survey of 1033 castor bean accessions from the United States Department of Agriculture (USDA), seed oil content ranged from 37.2 to 60.6% with an average of 48.2% (Wang et al. 2010). The variation range and sample distribution formed a normal distribution with 70% accessions having oil content between 45 and 51% (Wang et al. 2010). Impala Cycle 0 had a mean oil content of 50.33% which is comparable to that of most castor accessions. However, after one recurrent selection, Cycle 1 increased oil content to an average of 53.87% which falls into the top 2% of the entire USDA collection. Additional recurrent selection increased the mean oil content to 54.47% which put Cycle 2 at the top 1% of the USDA collection. These results indicate that single-seed recurrent selection is an effective method to increase the mean oil content in a castor bean population.

In our two selection cycles, we observed a quick drop of selection efficiency for oil content. The first cycle increased mean oil content from 50.33 to 53.87%, which represents about 7% mean improvement, whereas the second cycle increased from 53.87 to 54.47%, representing only 1% improvement. A wide range of selection responses of oil content to number of cycles or generations were reported among species with different genetic backgrounds and initial genetic variability. In maize, 54–70 genes were estimated to affect oil content (Dudley 2007). It took 110 cycles of recurrent selection to develop Illinois High Oil strain from mean oil 5–20% (Lucas et al. 2013). In oat, the starting background was an F_1 hybrid between a cultivar and a wild relative, which had large genetic variances and resulted in a linear increase of groat-oil content during nine cycles of selection (Frey and Holland 1999). Although the number of genes controlling castor oil content is unknown, the Impala base population is a cultivar rather than an F_1 hybrid, thus one possible reason for the lower rate of increase in the second cycle

could be a low initial genetic variability in the base population, which may have limited the genetic potential for further enhancement in oil content. A similar phenomenon was reported in soybean (Feng et al. 2004). Another possibility could be the existence of a biological upper limit of oil content in castor bean seeds. Among 1103 USDA castor germplasms, there are only two accessions found to have extremely high oil content (57.1 and 60.6%) (Wang et al. 2010). Given that the oil content gained in our Cycle 2 Impala population among the highest recorded in the USDA castor bean accessions, it is likely that our selection process has approached the genetic ceiling of improvement within the species.

4.1.3 Average Seed Weight Increased After Recurrent Selection

Because seed weight is an important yield component, it was of interest to know if the selection of seeds with higher oil content was associated with a change in seed weight in the populations generated by recurrent selection. As shown in Table 4.2, single-seed weight from base population (Cycle 0) ranged from 0.30 to 0.56 g with an average of 0.44 g (±0.06 g, SD) and the correlation between oil content and seed weight was moderate ($r = 0.43$, $p < 0.0001$). After selection, the seed weight range increased, showing 0.34–0.65 g in Cycle 1 and 0.34–0.71 g in Cycle 2 (Table 4.2 and Fig. 4.2) (Chen et al. 2016, 2017). The average seed weight and its correlation with oil content were also increased, showing 0.50 g (±0.06 g, SD) with $r = 0.63$ ($p < 0.0001$) in Cycle 1, and 0.54 g (±0.06 g, SD) with $r = 0.77$ ($p < 0.0001$) in Cycle 2 (Table 4.1, Fig. 4.2). The strong correlation after two cycles of selection indicated that seeds with higher oil content were frequently associated with heavier seeds.

Our results are similar to previous reports of castor bean cultivar BRS Energia, in which heavier seeds (0.45 g/seed) showed higher oil content (54.7%) compared with lighter seeds (0.25 g/seed, 48.2% oil content) (Severino et al. 2015). Detailed characterization of seed morphology indicated that in BRS Energia, when seeds were de-coated (leaving mainly endosperm tissue containing oil), there was little variation in average oil content between a light 0.25 g seed

Table 4.2 Cycle means, standard deviation (SD), minimum (min), maximum (max), and correlation (r) between oil content and weight

	Number of seeds	Seed weight (g)		r ($p < 0.0001$)
Population		Mean ± SD	Min–max (g)	
Cycle 0	233	0.44 ± 0.06	0.30–0.56	0.43
Cycle 1	770	0.50 ± 0.06	0.34–0.65	0.63
Cycle 2	1859	0.54 ± 0.06	0.34–0.71	0.77

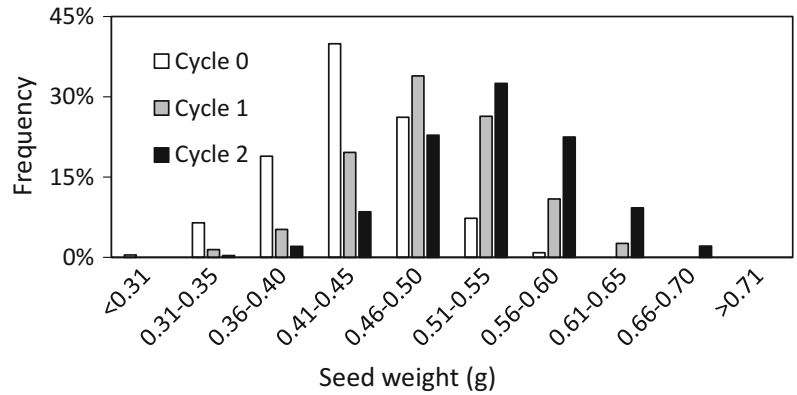

Fig. 4.2 Comparison of frequency distributions of seed weight in cycle 0, cycle 1, and cycle 2 of individual castor seeds

and heavy 0.45 g seed, estimated at 67 and 68.9%, respectively. In fact, the ratio of seed coat to seed weight in heavier seeds (or larger seeds) was reduced. Therefore, larger seeds contained proportionally less seed coat than smaller seeds. Furthermore, the heavier or larger seeds were associated with higher density in BRS Energia (Severino et al. 2015). Thus, it is likely that lighter seeds might hold internal empty spaces. The relationships among seed oil content, seed weight, seed coat, and endosperm characteristics in Impala remain to be investigated. Nonetheless, our findings suggest that it might be possible to further increase castor bean seed oil content by screening larger or heavier seeds. Despite of the strong correlation between seed oil content and weight observed both in Impala and BRS Energia, there was no significant correlation among 40 different castor bean genotypes ($r = 0.014$, $p = 0.49$) (Severino et al. 2015). A very weak correlation ($r = 0.157$, $p < 0.0001$) between seed weight and oil content was reported among 1033 USDA castor accessions (Wang et al. 2010) suggesting that strong positive correlations between seed oil content and weight can be found within the same castor bean cultivar.

4.2 Evaluation of Seeds from Field-Grown Populations

Recurrent selection in castor bean has also been tested in field trials performed at the University of California-Davis (Davis, CA) (DMS latitude: 38° 32′ 41.7″ N, longitude: 121° 44′ 25.9″ W). In these experiments, the Control seeds were the same batch of seeds as Cycle 0 base population of Impala as described above and the Test seeds were the Cycle 2 Impala population that had showed increased oil content in greenhouse settings. Impala is a dwarf cultivar and grows to ~1.5 m tall with a ~0.6 m spread. Each 400-seed plot was planted in five rows with spacing of 1.5 m between rows and 0.9 m between plants. Two control plots (Control A and Control B) and two test plots (Test A and Test B) were planted. The plots were spaced ten rows apart. The field was maintained by hand-weeding and surface irrigated with complete fertilizer. All seeds were hand-planted in the first week of June. Mature Impala plants produce monoecious inflorescences or racemes bearing both male and female flowers. To maintain the genetic purity of seeds to be produced, over 400 racemes (at least one from each plant) in each plot were covered with white cloth bags. Mature racemes from each plot were hand-harvested in late September and early October. The field trial was duplicated in two years (Year 1, 2011 and Year 2, 2012). Detailed weather information regarding 2011 and 2012 in Davis, CA, is documented at https://www.timeanddate.com/weather/usa/davis/historic?month=8&year=2011. In 2011, the daily mean temperatures were 21 °C in June, 23 °C from July to September, and 17 °C in October and in 2012, the daily mean temperatures were 22 °C in June, 24 °C in July and August, 22 °C in September, and 18 °C in October. No precipitation occurred from June to September in either year. Very light precipitation (<0.1 inch) occurred in October of both years. The general monthly weather climate in 2011 and 2012 was similar to the annual weather averages at Davis, CA (https://www.timeanddate.com/weather/usa/davis/climate).

Impala seedlings emerged from the soil one to two weeks after planting. Seedling germination rates in all plots and both years were high, at 98–100%. This indicates that selection of high-oilseeds does not affect seed viability. Flowering time, measured as days from planting to 50% of first raceme initiation and can vary among varieties. Some extra-early (26 days) and late varieties (120 days) are known (Anjani 2010). In each year, both Control and Test plants initiated their first racemes 60–74 days after planting, indicating that selection for oil content did not influence time of flowering. After flowering, both Control and Test plants took an additional 60–65 days for developing seeds to mature. During the entire season, no difference in growth pattern was noticed between control and test plants. Temperature influences seed maturation time (days from flowering or pollination to seed mature) (Severino and Auld 2014). The

observed 60–65 days to maturation of the most primary racemes of Control and Test plants occurred under a mean temperature from August to September of 23 °C in 2011 and 22–24 °C in 2012. Similar days to maturation were found in our previous study of strain PI215769 castor grown at a mean temperature of 23 °C (Chen et al. 2004). These results are consistent with a castor bean seed maturation study using a degree-days approach, which is a weather-based index used to predict number of days required for a crop to reach maturity, with the variety Brigham (Severino and Auld 2014).

There was no significant difference in the data obtained between Controls A and B, between Tests A and B, or between Years 1 and 2 (Table 4.3 and Figs. 4.3, 4.4) (Chen et al. 2018). Therefore, data from Control seeds or Test seeds are pooled for a succinct description here. The oil content of Control seeds ranged from a minimum of 30.13% dry weight to a maximum of 57.59% dry weight, with an average of 50.5% dry weight, similar to Cycle 0 seeds, which had an average of 50.3% dry weight (Table 4.1). Test seeds had oil contents from 30.13 to 59.41% dry weight, with an average of 54.01% dry weight, which is also comparable to the mean oil content of 54.47% dry weight in Cycle 2 seeds (Table 4.1). Single-seed weight from Control populations ranged from 0.31 to 0.58 g, with an average of 0.44 g. Test seeds had single-seed weights ranging from 0.31 to 0.64 g, with an average of 0.54 g. Thus, the average single-seed weight of Control and Test seeds remained the same as those of Cycle 0 (0.44 g) and Cycle 2 (0.54 g), respectively (Table 4.2). There were no significant differences in correlations between oil content and seed weight between Control seeds ($r = 0.43$ or 0.44, $p < 0.0001$) (Table 4.3) and Cycle 0 ($r = 0.43$, $p < 0.0001$) (Table 4.2), or between Test seeds ($r = 0.71$ or 0.72, $p < 0.0001$) (Table 4.3) and Cycle 2 ($r = 0.77$, $p < 0.0001$) (Table 4.2). In addition, the frequency distribution of Control seeds and Test seeds for oil content (Fig. 4.3) and seed weight (Fig. 4.4) resembled the corresponding Cycle 0 and Cycle 2 populations (Figs. 4.1, 4.2), respectively.

The results of the field test support our previous greenhouse study that recurrent selection through screening of single seeds is an effective method to improve seed oil content and thus seed weight in castor bean. Our studies also support the notion that seed oil concentration is a heritable trait in castor bean, as demonstrated in other crops (Bauman et al. 1963; Frey and Holland 1999; Feng et al. 2004; Lucas et al. 2013). A detailed study on castor-oil yield components (number of racemes, number of seeds per raceme, seed weight, and seed oil content) indicated that castor has adapted to environmental changes by adjusting the number of racemes and seeds per plant, but barely changed seed oil content and weight (Severino and Auld 2013).

Table 4.3 Average seed oil content, weight, standard deviation (SD), minimum (min), maximum (max), and correlation (r) between oil content and weight

	Population	n	Oil content (% dry weight)		Seed weight (g)		r ($p < 0.0001$)
			Mean ± SD	Min–max	Mean ± SD	Min–max	
Year 1	Control A	3027	50.57 ± 1.84	32.62–57.59	0.43 ± 0.05	0.33–0.58	0.43
	Control B	3104	50.36 ± 1.74	30.38–56.94	0.45 ± 0.05	0.31–0.57	0.44
	Test A	3115	53.95 ± 1.94	33.45–58.36	0.53 ± 0.06	0.34–0.62	0.71
	Test B	3231	54.07 ± 2.04	35.26–59.06	0.54 ± 0.04	0.31–0.64	0.72
Year 2	Control A	3054	50.77 ± 1.84	30.13–57.41	0.45 ± 0.04	0.33–0.55	0.44
	Control B	3206	50.45 ± 2.01	31.45–57.11	0.44 ± 0.05	0.32–0.56	0.43
	Test A	3124	53.88 ± 2.06	30.13–59.41	0.53 ± 0.05	0.34–0.61	0.71
	Test B	3087	54.12 ± 1.85	30.13–58.87	0.54 ± 0.06	0.33–0.63	0.71

n = number of seed

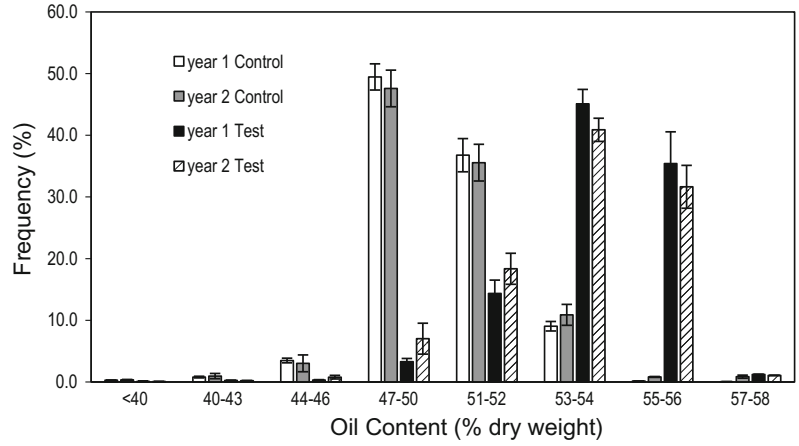

Fig. 4.3 Comparison of frequency distribution of seed oil content. All data are average of duplicates ±SD

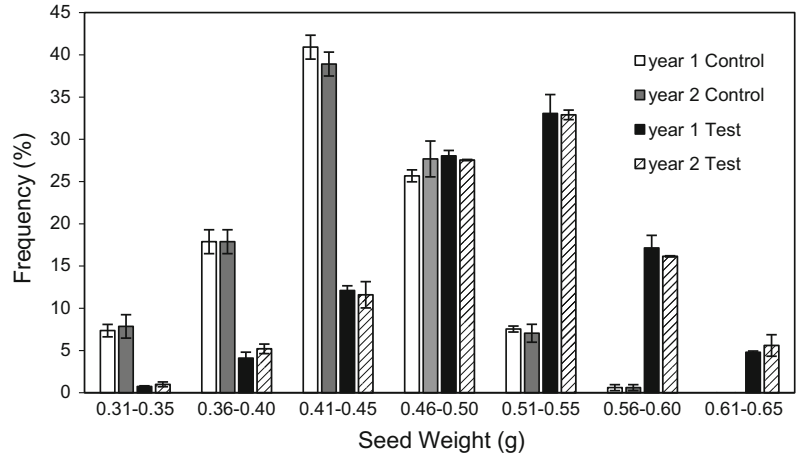

Fig. 4.4 Comparison of frequency distribution of single-seed weight. All data are average of duplicates ±SD

Castor bean seed oil content was not affected by planting density (Soratto et al. 2012), nor by growing the same population in different years (Anastasi et al. 2015). Our studies using greenhouse (Tables 4.1, 4.2 and Figs. 4.1, 4.2) and field settings (Table 4.3 and Figs. 4.3, 4.4) were conducted in different geographic locations but under similar water and fertilizer conditions. The temperature during castor bean growth in the field in these two years was also similar to the greenhouse temperature as described above. In addition, the studies used seeds mostly from primary racemes and were based on two consecutive years at one field location. Taken together, we observed no significant differences between field- and greenhouse-grown castor bean seed in oil content and weight. Seed oil and other storage compounds are important reserves for supporting seed germination and seedling growth and provide basic assurance for the next generation's viability. Negative side effects of recurrent selection have been reported in other crops. For example, increased oil content was accompanied by decreased oil yield in sweet corn (Rosulj et al. 2002) and grain yield in oat (Frey and Holland 1999). The seed and oil yields of this new Impala population still need to be investigated. As oil is the primary product with economic value in castor bean, development of a cultivar with enhanced oil content will provide a new genetic resource for increasing the oil yield of this important crop.

4.3 Conclusions

Recurrent selection through screening single seed is an effective method to improve oil content in castor bean. Two cycles of recurrent selection increased the mean oil content of Impala from 50.33 to 54.47%, reaching levels comparable to those at the top 1% of 1103 castor bean lines collected by USDA. As a consequent result, we found that seed weight was also increased after recurrent selection and the strong correlation uncovered between seed oil content and weight will allow further improvement of oil content by screening heavier or larger seeds in a population. Given that single-seed recurrent selection was successful under field conditions, this method should be applicable to commercial castor bean crops.

References

Anastasi U, Sortino O, Cosentino SL, Patanè C (2015) Seed yield and oil quality of perennial castor bean in a Mediterranean environment. Int J Plant Prod 9:99–116

Anjani K (2010) Extra-early maturing germplasm for utilization in castor improvement. Ind Crops Prod 31:139–144

Bauman LF, Conway TF, Watson SA (1963) Heritability of variations in oil content of individual corn kernels. Science 139:498–499

Caupin HJ (1997) Products from castor oil: past, present and future. In: Gunstone FD, Padley FB (eds) Lipid technologies and applications. Marcel Dekker, New York, pp 787–795

Chen GQ, He X, Liao LP, McKeon TA (2004) 2S albumin gene expression in castor plant (*Ricinus communis* L.). J Amer Oil Chem Soc 81:867–872

Chen GQ, Johnson K, Morales E, Mackey B, Lin J-T (2016) Rapid development of a castor cultivar with increased oil content. Ind Crops Prod 94:586–588

Chen GQ, Johnson K, Morales E, Mackey B, Lin J-T (2017) Corrigendum to "rapid development of a castor cultivar with increased oil content" [Ind. Crops Prod 94 (2016) 586–588]. Ind Crops Prod 101:103

Chen GQ, Johnson K, Morales E, Ibáñez AM, Lin J-T (2018) A high-oil castor cultivar developed through recurrent selection. Ind Crops Prod 111:8–10

Conway TF, Earle FR (1963) Nuclear magnetic resonance for determining oil content of seeds. J Amer Oil Chem Soc 40:265–268

Dudley JW (2007) From means to QTL: the illinois long-term selection experiment as a case study in quantitative genetics. Crop Sci 47:S20–S31

Feng L, Burton JW, Carter TE Jr, Pantalone VR (2004) Recurrent half-sib selection with testcross evaluation for increased oil content in soybean. Crop Sci 44:63–69

Frey KJ, Holland JB (1999) Nine cycles of recurrent selection for increased groat-oil content in oat. Crop Sci 39:1636–1641

Johnson W Jr (2007) Final report on the safety assessment of *Ricinus communis* (castor) seed oil, hydrogenated castor oil, glyceryl ricinoleate, glycerylricinoleate SE, ricinoleic acid, potassium ricinoleate, sodium ricinoleate, zincricinoleate, cetyl ricinoleate, ethyl ricinoleate, glycol ricinoleate, isopropylricinoleate, methyl ricinoleate, and octyldodecyl ricinoleate. Int J Toxicol 26:31–77

Lucas CJ, Zhao H, Schneerman M, Moose SP (2013) Genomic changes in response to 110 cycles of selection for seed protein and oil concentration in maize (chapter 12). In: Becraft PW (ed) Seed genomics. Wiley, London, pp 217–236. https://doi.org/10.1002/9781118525524.ch12

Mutlu H, Meier MAR (2010) Castor oil as a renewable resource for the chemical industry. Eur J Lipid Sci Technol 112:10–30

Rosulj M, Trifunovic S, Husic I (2002) Nine cycles of mass selection for increasing oil content in two maize (*Zea mays* L.) synthetics. Genet Mol Biol 25:449–461

Rubel G (1994) Simultaneous determination of oil and water contents in different oil seeds by pulsed nuclear magnetic resonance. J Amer Chem Soc 71:1057–1062

Severino LS, Auld DL (2013) Seed yield and yield components of castor influenced by irrigation. Ind Crops Prod 49:52–60

Severino LS, Auld DL (2014) Study on the effect of air temperature on seed development and determination of the base temperature for seed growth in castor (*Ricinus communis* L.). Aust J Crop Sci 8:290–295

Severino LS, Auld DL, Baldanzi M, Cândido MJD, Chen G, Crosby W, Tan D, He X, Lakshmamma P, Lavanya C, Machado OLT, Mielke T, Milani M, Miller TD, Morris JB, Morse SA, Navas AA, Soares DJ, Sofiatti V, Wang ML, Zanotto MD, Zieler H (2012) A review on the challenges for increased production of castor. Agron J 104:853–880

Severino LS, Mendes BSS, Lima GS (2015) Seed coat specific weight andendosperm composition define the oil content of castor seed. Ind Crops Prod 75:14–19

Sharma V, Kundu PP (2006) Addition polymers from natural oils—a review. Prog Polym Sci 31:983–1008

Soratto RP, Souza-Schlick GD, Fernandes AM, Zanotto MD, Crusciol CAC (2012) Narrow row spacing and high plant population to short height castor genotypes in two cropping seasons. Ind Crops Prod 35:244–249

Sprague GF, Miller PA, Brimhal B (1952) Additional studies of the relative effectiveness of two systems of selection for oil content of the corn kernel. Agron J 44:329–331

Wang ML, Morris JB, Pinnow DL, Davis J, Raymer P, Pederson GA (2010) A survey of the castor oil content, seed weight and seed-coat colour on the United States department of agriculture germplasm collection. Plant Genet Resour C 8:229–231

Wang ML, Morris JB, Tonnis B, Pinnow D, Davis J, Raymer P, Pederson GA (2011) Screening of the entire USDA castor germplasm collection for oil content and fatty acid composition for optimum biodiesel production. J Agric Food Chem 59:9250–9256

Molecular Linkage Maps in Castor Bean

5

Xuegui Yin, Jiannong Lu, Rukam Singh Tomar, Rajinder Singh Chauhan and Kwadwo Gyapong Agyenim-Boateng

Abstract

With the development of the global economy, the demand for castor bean and castor oil is increasing rapidly, but its cultivation suffers from a lack of high-yielding varieties due to limited genetic research. In this chapter, we describe novel SSR markers developed from the castor bean genome sequence and the first SSR-based genetic linkage map of castor bean, constructed with three different F_2 populations derived from crosses between the YC2, YF1, and YF2 lines. The SSR density in the castor bean genome is approximately 15.81 SSR/Mbp, and the frequency of SSR motifs decreased with the increase of repeat unit size. Dinucleotide and trinucleotide repeats, with (AT)n and (AAT)n are the most common repeat units. The linkage map consisted of 331 markers, distributed on ten linkage groups (LGs), encompassing 1164.73 cM, with an average marker interval of 3.63 cM. We will also discuss the first high-density genetic map of castor by using SLAF markers, developed by specific length amplified fragment sequencing (SLAF-seq). This map contains 4300 markers as well as 120 SSR markers with an average marker interval of 0.35 cM, making it the densest castor bean genetic map. These genetic resources are expected to facilitate castor bean research and breeding as well as comparative genomics analyses within the spurge family.

5.1 Introduction

Genetic mapping is the process of determining the order of and relative distance between genetic markers (specific DNA sequences or heritable elements that generate a detectable phenotype) on a given species' chromosomes. The relative location of genetic markers is determined based on their pattern of inheritance and is represented on a genetic or linkage map. The frequencies of recombination between markers due to the crossover of homologous chromosomes during meiosis are used to construct genetic maps. The genetic distance between markers in a genetic map is proportional to the recombination

X. Yin (✉) · J. Lu · K. G. Agyenim-Boateng
Agricultural College, Guangdong Ocean University, Zhanjiang, Guangdong, China
e-mail: yinxuegui@126.com

R. S. Tomar
Department of Biotechnology, Junagadh Agricultural University, Junagadh, Gujarat, India
e-mail: rukam@jau.in; rukamsingh@gmail.com

R. S. Chauhan
Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan, India

C. Kole and P. Rabinowicz (eds.), *The Castor Bean Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97280-0_5

frequency between them and does not necessarily correlate with the physical distance between markers on a chromosome. The genetic and physical order of genetic markers on a chromosome is the same, but distance between markers is not (Wei et al. 2007). Linkage maps have been constructed in many important species using different types of markers such as protein polymorphisms, restriction fragment length polymorphism (RFLP), random amplification of polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), sequence-related amplified polymorphism (SRAP), inter-simple sequence repeat (ISSR), simple sequence repeat, also called microsatellites (SSR), and single nucleotide polymorphism (SNP). Although a rather large variety of markers exist, each with advantages and drawbacks, microsatellites are very commonly used for linkage mapping (Goldstein and Schlotterer 1999). Genetic maps facilitate the identification of the genetic region of a trait of interest, providing a mechanism to track the cosegregation of genetic markers with traits in segregating populations. Such marker tracking can be used for the selection (marker-assisted selection) of individuals that harbor genes responsible for agronomically important traits and thus, serve as an aid in crop improvement (Morgante and Salamini 2003; Tuberosa and Salvi 2006). The comparative analysis of genetic maps can be used to understand the processes that led to the evolution and diversification of a species. Between related taxa, comparative mapping can reveal regions of chromosomal synteny (conservation of gene order). Within a taxon, mapping can pinpoint regions of chromosomal duplication derived from polyploidization events (Helentjaris et al. 1988; Moore et al. 1995; Bennetzen and Freeling 1997; Devos and Gale 1997; Feuillet and Keller 2002; Wei et al. 2007). High-density genetic maps have been shown to serve many purposes in both basic and applied researches. Such maps are the cornerstone of a variety of biological studies including map-based cloning, association genetics, and map-assisted breeding, especially for organisms that pose severe challenges for genomic sequencing due to their large and repetitive genomes (Wu et al. 2011). The construction of highly saturated maps is often seen as a time-consuming process, mainly if they employ different parental stocks and markers that are not easily transferable (Diaz et al. 2011). Merging different genetic maps increases marker density without additional genotyping, expands marker portability (i.e., polymorphic markers can be used in more than one population), improves marker alignment precision (i.e., congruent anchor maker position), and broadens inferential capabilities (i.e., cross-population prognostication) (Diaz et al. 2011). With further research, genetic map will be more perfect and play an increasingly important role in studies on castor genetics and breeding.

5.2 SSR Markers Developed in Castor Bean

SSR markers are codominant, highly reproducible, very abundant, and evenly distributed throughout the genome. Most SSR loci are defined by a unique pair of primers (Chen et al. 2006) and therefore, they are widely used in different genetic analyses, map construction, and QTL mapping. A limiting factor for SSR was the development cost in the past, but their value as markers made the investment worthwhile for their applicability in genome mapping studies (Rallo et al. 2000; Zane et al. 2002). Whole genome sequencing technologies have made SSR markers more affordable, facilitating their identification by computational genome mining (Cordeiro et al. 2002; McCouch et al. 2002; Song et al. 2010; Tan et al. 2014; Wei et al, 2014; Liu et al. 2016). High-density SSR-based genetic maps have been constructed in many crops, including rice (McCouch et al. 2002), wheat (Gupta et al. 2002; Song et al. 2005), maize (Sharopova et al. 2002), potato (Milbourne et al. 1998), and sesame (Rallo et al. 2000).

Until recently, large-scale development of castor bean generic markers had not been conducted. Foster et al. (2010) reported the development of 232 high-quality SNPs and their use to genotype a worldwide castor collection of 488

germplasm samples, resulting in the identification of five different worldwide genetic clusters. Only a few castor bean SSR markers had been developed before 2014, which severely hampered genetic studies. Allan et al. (2008) assessed the genetic diversity using AFLPs and SSRs, found the relatively low genetic diversity in castor bean germplasm and that there is no geographic structuring of genotypes. Additionally, SSRs were more informative than AFLPs as they showed higher polymorphism, heterozygosity, and range of genetic distances. Bajay et al. (2009, 2011) developed 23 SSR markers from microsatellite-enriched libraries and identified 38 genotype accessions from castor bean germplasm from the Brazilian Agricultural Research Company (EMBRAPA). The number of alleles observed for each locus ranged from two to five, with an average of 3.3 alleles per locus. The information derived from microsatellite markers led to significant gains in conserved allelic richness and provided support to the implementation of several molecular breeding strategies for castor bean. Seo et al. (2011) detected 28 SSR markers from microsatellite-enriched libraries. A total of 73 alleles were detected in 72 accessions, with an average of 3.18 alleles per locus, and the polymorphism information content (PIC) ranged from 0.03 to 0.47 (mean = 0.26). Values for observed (HO) and expected (HE) heterozygosity ranged from 0.00 to 0.19 (mean = 0.11) and from 0.04 to 0.54 (mean = 0.31), respectively. Machado and Silva (2013) developed microsatellite primers from castor bean sequences deposited at the GenBank/NCBI using the bioinformatic tools Websat and Net Primer. A total of 30 primer pairs were designed. The primer pairs presented an average guanine/cytosine (GC) percentage of 47.29% and amplified fragment sizes varying from 128 to 381 bp. Twenty-nine pairs of SSR primers (96.7%) were validated, of which nine were polymorphic (23.3%). Qiu et al. (2010) developed 118 polymorphic SSRs derived from expressed sequence tags (EST) which displayed moderate gene diversity with an average heterozygosity of 0.41 by screening 24 castor bean samples collected from different countries. A total of 350 alleles were identified from 118 polymorphic SSR loci, ranging from 2 to 6 per locus with an average of 2.97. This study also found that trinucleotides were the dominant motifs in EST-derived SSRs. Genetic relationships among 24 germplasms were investigated using the genotypes of 350 alleles, showing a geographic pattern of genotypes across genetic diversity centers of castor bean. Pranavi et al. (2011) reported 92 novel polymorphic EST-derived SSRs and used them in genetic purity assessment of hybrids. Sequence assembly and clustering of 57,895 ESTs of castor bean resulted in the identification of 10,960 unigenes (6459 singletons and 4501 contigs) having 7429 SSRs. On an average, the unigenes contained 1 SSR for every 1.23 kb of unigene sequence. The identified SSRs mostly consisted of dinucleotide (62.4%) and trinucleotide (33.5%) repeats. The AG class was the most common among the dinucleotide motifs (68.9%), whereas the AAG class (25.9%) was predominant among the trinucleotide motifs.

The release of the castor bean draft genome sequence that spans 352.27 Mb in nearly 26,000 scaffolds (or fragments of assembled sequence) (Chan et al. 2010) facilitated the development of SSR markers. Although the sequencing depth is only 4.6-fold, it still provided substantial information for marker development. By mining the castor bean genome, Tan et al. (2014) developed 1435 SSR primer pairs, among which 670 (46.7%) were polymorphic between six accessions. In total, 18,647 SSR loci were identified, with a density of one SSR per 18.89 kb. Dinucleotide repeats were the most frequently observed microsatellites, although the AAT repeat motif was also prevalent. Trinucleotide motif loci contained a higher proportion of polymorphisms (48.5%) than dinucleotide motif loci (39.2%). Polymorphism levels were positively correlated with the increasing number of repeat units in the microsatellites. These SSRs were also used to evaluate the phylogenetic relationship among 32 castor bean varieties. Cultivars developed at the same institute clustered together, suggesting that these cultivars have a narrow genetic background.

Another 3000 SSR primer pairs were developed by Liu et al. (2016) using the castor bean draft genome sequence. Using microsatellite Identification Tool Software (MISA) (http://pgrc.ipk-gatersleben.de/misa/), 25,828 sequence assemblies were mined for SSR motifs including dinucleotide repeats (DNRs), trinucleotide repeats (TNRs), tetranucleotide repeats (TTRs), pentanucleotide repeats (PNRs), hexanucleotide repeats (HNRs), and heptanucleotide repeats (HTRs) with 20, 10, 7, 7, 7, and 7 repeat units, respectively. Primer pairs were designed based on the flanking sequences of the detected SSR motifs, and the parameters were set as follows, primer length of 15–27 bp, amplification product size of 100–400 bp, melting temperature at 55–60 °C and GC content of 40–60%. As a result, 5546 potential SSR loci were identified, with a density of approximately one locus per 63.2 kb. Among these SSR loci, 108 were compound (more than one SSR in the same locus) and 98 of them were had two immediately adjacent SSRs, six with three SSRs, three with four SSRs, and one with five SSRs. The frequency of SSR motifs decreased as the repeat unit size increased (Tables 5.1 and 5.2). The motifs could be divided into 143 repeat types, containing 3618 DNRs, 1651 TNRs, 136 TTRs, 27 PNRs, 23 HNRs, and 91 HTRs, indicating a nonrandom distribution of SSR unit size in the castor bean genome. Although the types of DNRs and TNRs were only 10 and 35, respectively, they accounted for 65.25 and 29.77% of all SSR loci, respectively, with a density of one SSR per 96.91 and 212.37 kb, respectively.

Although SSR motifs larger than TNR were the most diverse (98 types in total), their frequency was low, with only 278 loci (4.99%) detected, representing a density of one SSR per 1265.82kb. The frequency of different motifs was very variable, even within the same-unit-sized SSRs. For example, AT was predominant in DNRs, followed by TA and AG,while AAT was common in TNRs (Table 5.2). The longest dinucleotide repeat locus was an AT motif that spanned 126 bp, and the longest trinucleotide repeat was an ATT motif with a length of 144 bp.

Based on 25,828 assemblies, the density of SSR distribution observed by Liu et al. (2016) was approximately 15.81 SSRs/Mbp which was lower than that observed in previous work by Tan et al. (2014), which reported a density of 52.9 SSRs/Mbp. The discrepancies may be due to a difference in the SSR detection criteria and primer design parameters used by the two studies. Also, each study used different versions of the genome assemblies, which may have accounted for some of the differences.

The polymorphic rate of SSR markers reported by Liu et al. (2016) was 27.38 and 59.6% when detected by limited biparent materials and ten randomly selected accessions, respectively, suggesting that diversity was lower in the biparent population than in the natural population. Regardless of the discrepancies, both studies by Tan et al. (2014) and Liu et al. (2016) substantially increased the number of castor bean SSR markers available, representing powerful tools for castor bean genetic research and breeding.

Table 5.1 Characterization of castor bean microsatellite markers

Repeat type	Counts	Total	Proportion to all SSR loci (%)	Average SSR distribution (1/kb)
DNR	10	3618	65.24	96.91
TNR	35	1651	29.77	212.37
TTR	38	136	2.45	2578.09
PNR	16	27	0.49	12,985.94
HNR	21	24	0.41	14,609.18
HTR	23	91	1.64	3582.98
Total	143	5546	100.00	

Table 5.2 SSR motifs identified in the castor bean genome sequence

SSR motifs	Number of repeats																									
	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	>30	Total
AC/GT														6	3	4	5	6	2	1	3			1	2	33
AG/CT														89	69	41	45	32	23	12	14	7	14	6	20	372
AT/AT														338	300	322	290	245	204	207	157	151	165	126	708	3213
AAC/GTT				2	1	2		2					1												0	8
AAG/CTT				129	82	36	23	12	4	4	1	3		1			1							1	1	298
AAT/ATT				297	231	179	134	107	80	59	43	41	33	20	19	11	12	10	6	6	1	1			6	1296
ACC/GGT				3	3	2																			0	8
ACT/AGT					1																				0	1
AGC/CTG				2	2																				0	4
AGG/CCT				6	3	1																			0	10
ATC/ATG				9	8	5	3	1																	0	26
AAAC/GTTT	2																								0	2
AAAG/CTTT	29	7	2																						0	38
AAAT/ATTT	46	14	4																						0	64
Other TTRs	20	5	5			1																			1	32
AAAAT/ATTTT	11																								0	11
Other PNRs	7	8	1																						0	16
HNRs	14	2	4	3																					0	23
HTRs	9	8	5	4	2	1	2	3	1	1	3	3	1		3	1		2		3	2	2	1	2	32	91
Total	138	44	21	455	333	227	162	125	85	64	47	47	35	454	394	379	353	295	235	229	177	161	181	135	770	5546

Tomar et al. (2017) downloaded the draft genome of 365 Mb with 35,370 scaffolds from NCBI (http://www.ncbi.nlm.nih.gov/Traces/wgs/?val=AASG02#contigs) for SSR mining again and identified 120,895 SSRs. The SSR containing 4143 sequences were used for the designing of primers using Primer3, and a total number of 662 primer pairs were identified.

5.3 The Genetic Map of Castor Bean

The first genetic map of castor bean was constructed in 2016 (Liu et al. 2016), and it was mainly based on SSR markers derived from genomic sequences. The map was built using three different populations derived from the cross combinations between the pistillate inbred line YC2 and the monoecious inbred lines YF1 and YF2, respectively, named population $F_{2\text{-}1}$, $F_{2\text{-}2}$, and $F_{2\text{-}3}$, each with 162, 182, and 161 individuals. Population $F_{2\text{-}1}$ came from the only pistillate F_1 individual (self-cross with artificially induced male flowers) of the YC2 × YF1 cross combination, and population $F_{2\text{-}2}$ came from one of the monoecious F_1 individuals of the same cross combination. The third F_2 population, $F_{2\text{-}3}$, was derived from the cross YC2 × YF2, which produced no pistillate F_1 individuals. The pistillate line YC2 was derived from the self-progeny of hybrid variety Jinbi No. 2, bred in Shanxi, China, and the monoecious lines YF1 and YF2 were derived from the self-progeny of the hybrid variety CSR6.181, bred in Costa Rica. The genetic distance (calculated using 30 important traits and 19 SRAP primers) between YC2 and YF1, and between YC2 and YF2 were 14.88 and 15.12, respectively, medium-to-large among 24 alternative key parents (the genetic distance between them ranged from 7.49 to 17.36), which was considered beneficial due to its potential to enhance heterosis. Moreover, there were sharp differences between each parent's phenotypic characteristics, spanning a wide range of important agronomic and morphological traits. These phenotypic differences included sex type (pistillate/monecious), plant height (220.4 cm/81.5 cm/80.6 cm), internode number in the main stem (11/31/30), stem and petiole color (green/red), resistance to castor blight (resistance/susceptible), leaf style (smooth and flat/wrinkled and cupped), spike type (column/pyramid), seed coat color (solid black/decorated pattern), wax on capsule and main stem (present/absent).

The markers used to construct the castor bean genetic map included SSR markers developed by Liu et al. (2016) plus nine SSR markers developed by Seo et al. (2011). To reduce the cost of map construction, some SRAP, SSRAP, ISSR, and three morphological markers were also used. SSRAP markers are a novel kind of marker invented by the author. SSRAPs are designed by randomly combining primer pairs between SSR and SRAP primer pairs, which proved to be an economical and efficient method for developing new markers using existing markers.

Three linkage maps were constructed with different populations using the program JoinMap 4.0 (Van Ooijen 2006), and an integrated linkage map with 331 markers was obtained (Table 5.3). The linkage map was composed of ten linkage groups (LGs) and spanned 1164.73 cM, with an average marker interval of 3.63 cM (Fig. 5.1; Table 5.3). On average, each LG covered 116.5 cM and contained 33.1 markers. Four large gaps of 58.68, 36.76, 29.38, and 22.99 cM remain in the map and therefore, more genetic markers still need to be added to saturate the map (Table 5.4).

Tomar et al. (2017) constructed another genetic map of castor bean with a F_2 population, containing 261 markers (76 RAPDs, 34 ISSRs, and 151 SSRs) assigned to 10 LGs corresponding to the castor bean chromosomes (Tables 5.5 and 5.6; Fig. 5.2). Assignment of markers to LGs was carried out using a LOD score value of 6 and recombination fraction of 0.3. The total map length was 1833.4 cM with an average marker interval of 6.93 cM. The length of LGs ranged from 121.5 cM (LG10) to 278.7 cM (LG9). The maximum average marker interval was found on LG3 (8.23 cM) and minimum in LG10 (5.78 cM). LG4 included the largest number of markers (11 RAPDs, 8 ISSRs, and 16 SSRs, representing 35 markers in total) with an average density of 0.15 markers/cM, while LG8

Table 5.3 Characteristics of the genetic linkage map of *Ricinus communis* L.

LG	LG size (cM)	Marker number	Average interval (cM)	Max gap	All marker percentage	Segregation distortion marker number	Segregation distortion marker percentage	Frequency of segregation distortion marker (%)
LG 1	131.09	38	3.54	11.14	11.48	5	9.43	13.16
LG 2	115.95	44	2.70	15.00	13.29	4	7.55	9.09
LG 3	162.11	51	3.24	58.68	15.41	8	15.09	15.69
LG 4	88.63	53	1.70	8.52	16.01	7	13.21	13.21
LG 5	107.22	30	3.70	18.55	9.06	7	13.21	23.33
LG 6	101.08	6	20.22	29.38	1.81	1	1.89	16.67
LG 7	140.96	30	4.86	22.99	9.06	2	3.77	6.67
LG 8	141.05	25	5.88	36.76	7.55	3	5.66	12.00
LG 9	95.07	39	2.50	15.61	11.78	10	18.87	25.64
LG 10	81.59	15	5.83	12.75	4.53	6	11.32	40.00
Total	1164.73	331	3.63	58.68	100.00	53	100.00	16.01

contained the lowest number markers (8 RAPDs, 1 ISSR, and 10 SSRs, 19 markers in total) with average density of 0.15 markers/cM.

5.4 Building the First High-Density Genetic Map in Castor

Xuegui Yin et al. (unpublished) constructed a high-density genetic map in castor bean using specific length amplified fragment (SLAF) sequencing in 2017. The F_2 mapping population consisted of 150 individuals derived from a cross between YC2 and YF2. SLAF library construction, high-throughput sequencing, sequence data grouping and genotyping were performed according to the procedure reported by Sun et al. (2013). A total of 194.63 Mb of sequence were generated using 200 bp paired-end reads and submitted to the NCBI SRA database under project accession number PRJNA414256. 89.83% of the produced sequence data were high quality and the GC content was 37.51%. 117,025 and 146,843 SLAFs were developed in female and male parents (Table 5.7), with an average coverage of 21.73-fold and 30.96-fold, respectively. On average, 108,686 SLAFs were developed in the offspring, with a 6.42-fold coverage (Table 5.7; Fig. 5.3). Of the 18,188 polymorphic SLAF markers identified, 4300 satisfied the following criteria: the sequence depth in the parents should be more than tenfold, marker missing rate should less than 0.3 in the F_2 population, and the *P*-value of segregation distortion should be less than 0.05. Those 4300 high-quality markers were used to construct the genetic map.

All of the 4300 SLAF markers plus 120 SSR markers were anchored onto 10 LGs. The final genetic map spanned a total of 1547.41 cM with an average marker interval 0.35 cM (Fig. 5.4).

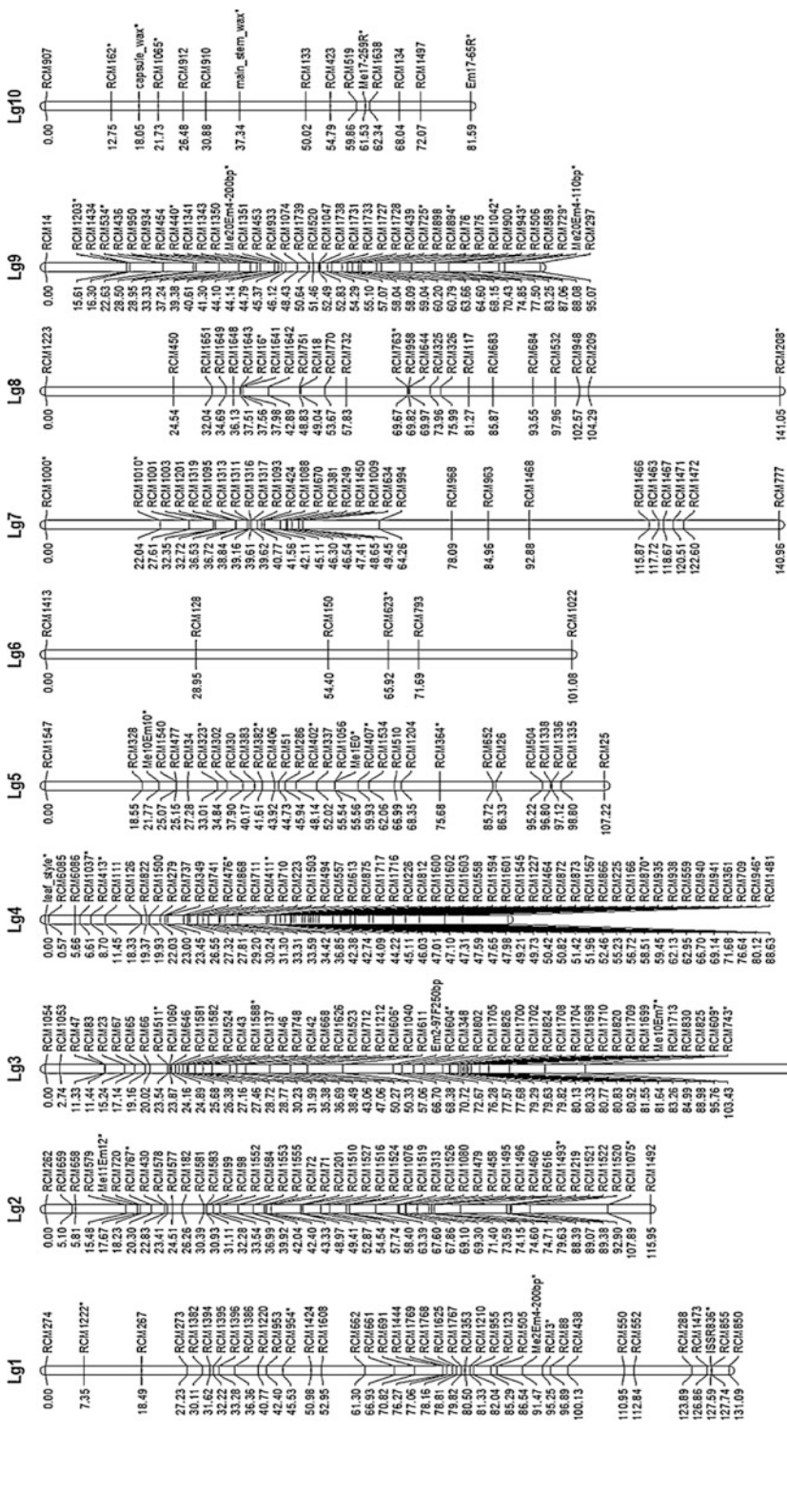

Fig. 5.1 Genetic linkage map of *Ricinus communis* L. The LG refers to linkage group

Table 5.4 Information of SSR markers located on the genetic map

Marker	ID	SSR motif	Forward primer (5′–3′)	Tm (°C)	Reverse primer (5′–3′)	Tm (°C)	Expected amplicon size (bp)
RCM3	27383	(AT)28	TTCTGAAGGCAAAGTTCTCG	58.2	TGTCAGCTTGAACAACGCA	60.2	316
RCM14	27394	(AT)34	AGTTTTCTGCCTTGATGGCT	58.9	ATACGATATGCCATGTGAAAAA	57.1	176
RCM16	27394	(TA)21	TGCACCGGTAAAAGAAGAGTT	58.9	CTGGATCTCGATAATCGGTG	58.1	381
RCM18	27395	(TTA)12	GGGCTTTTGGATTTTCCG	60.4	TGGGGCTAATGAGATTGAGTG	60.1	164
RCM23	27865	(TA)20	GGAAATGAAGGTGGTGGAGA	59.9	TGGCAAAATTGAGTCGCA	59.9	307
RCM25	27866	(AT)20	TCTCTCTCTCCCTCTCTCTCTCTC	59.1	AAATCCCTTTTGAGTTATGTAATAATG	57.4	209
RCM26	27866	(ATA)16	CTGCTCTCCCGTCACAAAA	60.0	CCGGTTAATCAATCAATACCTT	57.1	355
RCM30	28829	(TA)33	CCATGCTTGAAAATTTGGC	59.1	GACCCCTTAATGGAAGGCA	59.9	336
RCM34	28830	(ATT)14	TTTCTCACCCTATTGGATAACG	58.1	ATAGCCCATCGCCAAACA	60.0	330
RCM42	28842	(AAT)14	CTCTCATTAACGGAGCGTG	57.0	CGCCGCAAAATTAGAAGG	59.4	336
RCM43	28842	(AT)22	TCGAAGGGTCTAGAGATCGG	59.4	CATGTAGGCTCAATGGGGA	59.5	208
RCM46	28842	(TG)22	TTTCAGCGCAGCATATCG	59.6	TTGTCGTTGTGGGCATGT	59.5	370
RCM47	28842	(AT)25	GCTTTGGTGCTTTTGGGA	59.8	TGTGTCAAATACACATGTCAAAAT	57.0	281
RCM51	27889	(TA)33	TTGGATTATTCTGGTTATCACCTTT	59.2	CAGGTTCCGAAGAGCAGG	59.5	400
RCM65	27894	(TC)24	AGCCATAATGCAGGCTCG	59.9	TCCAGGGCCACACTTCAT	60.0	295
RCM66	27894	(TA)37	GCAAGCCGAGCCTGAATA	60.1	CAAGCTTGGATAATTTCTCTCTCTC	59.1	225
RCM67	27894	(ATT)12	CCCAATTAAATGGAAATCTTTTCT	58.9	TTGCACCTTCCTGCCATT	60.2	302
RCM71	27895	(TA)26	CCAAAGTTTTTGAGGGCCA	61.0	CGCCTAAGGAGTTCCCTTT	58.4	242
RCM72	27895	(AT)22	TTTTGAAACACTAAAATACCCCA	57.7	TTGATCCAAATGCAATTCTCA	59.1	310
RCM75	28863	(AAT)25	GAAGCCGACGCCAGTTAC	59.4	CATCATCATCACAATCCTTATTATCAC	59.9	207
RCM76	28863	(TTA)26	CTCCGCGGTATGACAAGG	60.2	CGAACGTAAATCCGGCTA	57.3	292
RCM83	27904	(TA)38	GCTGGTTTGGTACGGTTCTC	59.6	CCGCGTTTTCTAACTTTGTG	58.5	400
RCM88	28872	(TA)38	CAAGTGGAGCGTTATGGGA	59.7	GAAAGCAAGTTGATTATAGGCGA	59.8	206
RCM97	27914	(TA)27	AGCAGCTCATGCGTCTTAATC	59.6	TGGGAAAATCTTTAGTGCTATGC	59.7	391
RCM98	27914	(AAT)18	CACTCCCATTACGCTCTACCA	60.1	GTGCCTGCTACTCCCAGC	59.5	353

(continued)

Table 5.4 (continued)

Marker	ID	SSR motif	Forward primer (5′–3′)	Tm (°C)	Reverse primer (5′–3′)	Tm (°C)	Expected amplicon size (bp)
RCM99	27914	(AAT)13	CTGTGGGTCCAATTTCGC	60.0	GAGGTCACAATGGGCGAC	60.1	393
RCM111	27917	(TA)28	CACATGCGTTGCACGTAA	58.8	TTTTTAATGGTGTAGGTGAATCTTAAT	57.8	286
RCM117	27934	(CT)27	AACCCCACCGTGTGTTTC	59.2	AGAACCCAAAGCCCAAGC	60.2	241
RCM123	27945	(TA)39	TGGTTCCAAGGGACAACAA	59.9	GCCACGAGAGCTTGGAAA	60.1	365
RCM126	27955	(AGA)11	AACAAGCAACACAAAATGCAA	59.3	TTGTCCTGTGTTGTTGGACC	59.4	248
RCM128	28927	(ATA)17	CACCTTGATATGGTTGATTTATGA	58.0	GCACACACCAACCCACAC	59.4	191
RCM133	27956	(TCT)10	GCAGCAGCAGCAGCAGTA	60.2	GGCGTTTACATTGGGTGC	59.9	224
RCM134	27956	(AT)27	ACTTCTCGGGAGGGATCG	60.2	AGCTCGCCGTGCATGTAT	60.4	394
RCM137	27961	(TTAA)7	TTGTCCCCTTTGTTGATGC	59.5	AGGCCCAAGCTATTGGAAA	59.7	305
RCM150	27985	(TAA)12	CCCAGTAATTGGGACGCA	60.5	AAAAAGATGTCGTGGATGGA	58.0	237
RCM162	27996	(TA)23	TTTACCAAATTGCTCCCTAAAAA	59.1	CAAAAGACGTGCATGTGAAAA	59.8	253
RCM166	28006	(TAA)17	AACATTGTTTAACGCATGGAT	57.1	CAACTCAATTCTATTTGTGAAAGTG	57.2	141
RCM182	29005	(AT)28	CAAGACTTATTTTCAGCCACGTT	59.7	CCCTTCATTGAATCCTTGC	58.1	365
RCM201	29032	(AAT)11	GAACTAGGCAAAAAGTGGAGGA	59.8	CATGGATGAATAACCCAAAGC	59.3	262
RCM208	28076	(TA)21	TCCCCTCTTTTGGGTTCC	59.8	TGAGGCCTGGAAAGGAAG	58.9	292
RCM209	28076	(TTA)12	TTTCGTTTTTCCAATAAAGTCCA	59.9	GGAGCTCGCTATTAGGTTCG	59.1	373
RCM219	28093	(AT)36	TCAAATCCGCCAGAGAAAA	59.3	TGCTATCGGAATAGATCCTTCAA	60.1	394
RCM223	28094	(TAT)15	CAATCTGCAGCTTATTGGGA	58.9	TGAGGGGCCAAAAGAGTG	59.8	327
RCM225	28094	(GAT)12	TGGCCAAAACAGCAGATG	59.3	GCAGCTGAGTACCATCCCA	59.8	385
RCM226	28098	(TA)40	TGTTATGCTTGGAATTCTTTCA	57.5	CCAACAACAACTTCATTGCC	59.0	384
RCM249	28128	(TCT)14	CACCGAAAACACCATCCC	59.7	GCTGCTGCTGCCATTCTT	60.3	242
RCM259	28134	(AT)32	TCGAGCTTTTATGAAGGTGTGA	59.9	CATGTTCAGCAATCCATGC	58.6	217
RCM262	28140	(AG)23	CCAAGCTTCAGCCTCCAA	60.1	AAGGGCTTTGCGGAACAT	60.6	321
RCM267	28152	(CTT)10	CCAACAAGGCTGCCAATC	60.2	TGGAGGACCCAGTGGAGA	60.2	208

(continued)

Table 5.4 (continued)

Marker	ID	SSR motif	Forward primer (5′–3′)	Tm (°C)	Reverse primer (5′–3′)	Tm (°C)	Expected amplicon size (bp)
RCM273	28152	(TA)30	TTTTTCCCCACCCATTATTCT	59.5	ACGGCAGCATGCTTGAAT	60.4	190
RCM274	28152	(AT)23	GCCGGATAGTAATATTTTTGCG	59.9	TGAAGAACATGGACTTGGTCA	59.1	345
RCM279	29130	(AT)25	AAAAATGACTTTCATGCTTTCAA	58.0	AGGTAGGGGCCTGTTCCA	60.5	339
RCM286	28161	(TTA)10	AGCTGGTGTTGACAGGGG	59.7	GCTTGTCCCGACACTCTCA	60.0	236
RCM288	28162	(AG)21	ACGGCATTTTGAACGGAA	60.0	TCTCGATGCGATCCAGGT	60.3	328
RCM297	28170	(TA)23	TGTGCGTATTGACCCGTG	60.1	TCTTACAGTGGAAACTCATGAAAA	58.0	323
RCM302	28179	(TA)22	TGAGGAAAAGATTAAGGGCCTA	59.2	TTTCGCGCCGATTTTATC	59.7	384
RCM313	28200	(AG)28	ATGAGTGGGCGTGCTTTC	59.8	TGGTTCTGAAAGGGCCAA	60.2	201
RCM323	29169	(AT)20	TTGCAGTGACGACTGGGA	59.9	TGTACGAGGAAATGGCGTC	59.7	252
RCM325	29171	(TTA)22	GGATTTAAATGCAACTCTTTTGATG	60.2	CACACATGGTCTTGATGCG	59.7	370
RCM326	29171	(AG)25	ACCGACAAGCAGTCGAGC	60.1	GAGGCGAGGCAATTTTGA	59.9	326
RCM328	29172	(TA)25	GAGTGAAAGAGTTTGGACCCAC	60.0	CATTTGCTCTTTGACGGGA	59.8	379
RCM337	28225	(TTA)11 (TTG)11	TTTTTAATCACACTCATCACTCATTTT	59.4	CCAAAACCCTTCGCTCAA	59.8	352
RCM348	29200	(AT)22	TGTTTATGCATAGCACTTCCTTTT	59.3	TTGGTCCTCAACCCAAACA	59.9	245
RCM349	28238	(ATT)14	TGAAATCATCATAATCACACTCCTC	59.4	TGGATTCGGTAAATGGATACC	58.6	219
RCM353	29212	(AT)21	GCTTCTGGAGGGTCTTTGG	59.8	GCCCTATATTTCGGAGTTCCTT	59.8	195
RCM361	29216	(TA)30	TTTTCCTCCCCCAAAGGA	60.4	GGTGGGGTCGTGACAGTT	59.3	357
RCM364	28256	(AT)23	CCCTTAGCTGCAACGCAT	60.0	AGTTTTCAGGGGCAGTCG	58.8	386
RCM381	28266	(AAT)13	CCATCTCATCCATTGCTGG	60.0	TTTGGAGAGATTTAAATTCGGA	57.9	260
RCM382	29235	(TA)23	TGCTGCATCAAATTTCTGAAC	58.9	TCTAGCCCCATTCTCAGGC	60.3	323
RCM383	29235	(TCT)10	CCACTTCTGCACGCCATA	59.4	AGGCAGGTGCGTATCAGG	59.8	236
RCM402	28333	(TA)24	CATATTGTGTTCAACTTTTGTGAAA	58.2	TGAGGATCCTGCACACCA	59.7	156
RCM406	28333	(AT)25	CTTTTTCCTGGGCCTTTCT	58.4	TGGGTTGGAAAAGAATCCA	58.9	360
RCM407	28336	(AT)20	TGCATTCTTCTAAGCTATTATCCAA	58.6	CAGGAATGTTTTCCTTAGGTGG	59.9	381

(continued)

Table 5.4 (continued)

Marker	ID	SSR motif	Forward primer (5′–3′)	Tm (°C)	Reverse primer (5′–3′)	Tm (°C)	Expected amplicon size (bp)
RCM411	28345	(TA)31	TGGCAACCGATGAGATGA	59.7	AAATACACATGTTAAAGCACAAATCTT	58.7	338
RCM413	28345	(AT)22	TCAATCAAACATCAACTATTCTCATC	58.6	ACTGGTAATAATGGTCATTGCAT	57.5	235
RCM423	28357	(TC)26	TCATCATTGGCCCAACCT	59.9	TGATGGGAAAACGTTTCTTTG	60.0	223
RCM424	28312	(TA)20	TCACGCCTCTCTCTCTCTCTC	59.0	AGTCTCTCCGCGGCTCTT	60.2	373
RCM430	28392	(AG)29	AATGCAAAGGGGTGCAGA	60.2	GGAAGGCTTTGAGATTTGAGA	58.5	397
RCM436	28402	(TA)20	AAGATGTGGTTGTTTTTCAGTTATTG	59.8	TCCTCAAAAGCGATAAGCG	59.1	359
RCM438	27428	(TA)28	TCATCCAAAGAAAGAGAAATTCAA	59.3	TCCTCCTTGTGTTAAACCAAAGTT	60.3	340
RCM439	28421	(TA)27	AGCGCATAGGCTGGAAAA	59.5	AGTTCAGCCTCCATCCCC	60.0	284
RCM440	28421	(AT)21	GCGGGAATACTCCCAGTG	59.0	GCCACATGCTTTCACCTT	57.1	400
RCM450	27446	(GA)26	GGCATGTTTCGAGGTGGT	59.5	ACCCTTCTTCGCGTGCTA	59.5	218
RCM453	28429	(TA)23	AAAGCACATATTTCCGCCA	59.1	GCTTGACTCCATGGTCCG	60.2	347
RCM454	28429	(AT)20	TACGCTCGATACTCCGCC	59.9	ATCCTCACCCCCACCATC	60.6	363
RCM458	27467	(TAA)11	GGCCAAACCAATGTGCTC	60.1	CCGCCACAACTGCCTAAC	60.3	215
RCM460	27467	(TA)22	AATTGGTGAGAATCCATTTGC	58.9	TGGAGATTCAACCCCAACA	59.9	341
RCM464	28448	(AAT)28	CCAGGTCATGACGAGCAG	58.8	CGGCAGCCTCCTAAGGTT	60.4	357
RCM476	28449	(AG)24	GGGGAGTATGACATCGAGC	58.0	TGGATTGATGGATTGGGTTT	60.0	395
RCM477	28452	(ATA)20	CACCGACCACCGACCTAC	60.0	CCTTTCAGTCATATTAGAACACTCG	58.5	382
RCM479	28455	(TTA)19	CCGACACATTCTAGATCAATTCC	59.8	TGGAGAAGGAGGGTATCGTG	60.1	202
RCM494	28470	(TA)25	TGAAGGATGTTCATTATTTATGATTT	57.3	TGTTTCAATGGGCATCACA	59.5	215
RCM504	27504	(TA)20	TGCCAATTCCAGCATTGTT	60.1	CAAGAAGCTGCTGTGCCTT	59.3	308
RCM505	28490	(AC)20	TTGCGGAGAGAATCCAGC	60.0	CCAATCCCACAATGCAGA	58.9	320
RCM506	27506	(AT)23	TGACTTCTAGTTGCCCATGTTT	58.8	AGGATCCTCCTCTTTGCCA	59.7	334
RCM510	28492	(ATA)11 (AATA)7[a]	TGCTTAATTGAATGCGTGGA	60.2	GGCATAAAATGGGCGTTG	59.9	256
RCM511	28495	(TA)20	TCCAAACTGCAACCCCAT	59.9	GGGAGTGAAGGGGAAGGA	60.0	278

(continued)

Table 5.4 (continued)

Marker	ID	SSR motif	Forward primer (5′–3′)	Tm (°C)	Reverse primer (5′–3′)	Tm (°C)	Expected amplicon size (bp)
RCM519	27524	(AT)20	CCGTTGAGGATGAATATGGC	60.3	TTGTATGCGCGCTTTTCC	60.9	178
RCM520	28515	(TTC)29	GGGAATCACCGAGATCCA	59.3	TGCTAATGGCATCCAAAACA	60.1	384
RCM523	27533	(AT)21	TCCCCCTTTGACAATTTCC	59.7	ACCCATTTTCATTGTCATTCA	57.8	306
RCM524	27533	(AC)25	TGGGGAGAAATGGCAGAA	60.1	ATGAGAGGGAAGAGGCGG	60.3	345
RCM532	27538	(CT)22	TCCCATTAAAATTGTCTCCCC	60.0	TGGGACTTGGGCTCTTGA	60.3	367
RCM534	28533	(AT)33	TGCGAAGTCCTTCTTCTCG	59.3	TTTTGAGGCGCATGTTTAAT	58.3	257
RCM550	27568	(AT)21	TTTTTGGACTCGTGCACTTT	58.4	CCAAGCCACCAGATTTGC	60.2	377
RCM552	27568	(TA)21	TCCATGAGTAATTTGTCGTTTGA	59.5	TGTGTTGTCGTGTGACATGAAT	59.9	398
RCM557	27574	(TTA)16	AATCGGGCTCTGGGCTAC	60.2	TTTCTGTTTCAAACGTTCATTTTT	59.1	155
RCM558	27574	(TA)23	CAAGCATGGAATCAATTTCTCA	60.1	GATGCACATTTTTGGCCG	61.0	365
RCM559	27574	(AAT)15 (ATT)13	TTCCAATGATTTTTGGCTCA	59.1	AACATGTGGCCAGTCACTTT	58.1	266
RCM577	28609	(TAA)11	TGCTGATGTGGATCGCTG	60.5	GGGAAAATGCTGTCGGTG	60.0	324
RCM578	28609	(AT)22	GGAAGGGGACCATTAGGC	59.3	TGATGCAGTTCTAAATGCTCAGA	60.0	325
RCM579	28609	(AT)37	TGTCAAAGATAAATGCACATATCAAA	59.8	AACACGTCAGCTAAACAATAACATT	58.3	265
RCM581	28611	(TTA)15	TTGAGAGAAAATTTGAGCGTGA	60.0	CCGTACTTGTAAGACCCGAGA	59.2	344
RCM583	28611	(AT)21	CCCTCCATCCTCCATTCC	60.2	TGTGCCAACCCCCATAGT	59.8	251
RCM584	28611	(TA)35	TCAAAATTCATAACCAAGGCA	58.1	TGCCTATTAAGATACCAACCTTTGA	60.2	382
RCM589	28613	(TA)21	TGGACAGAGATTTAAGCGGAA	59.8	AATGTTTCATTTAATTGATTAGGAGAT	57.1	400
RCM604	27613	(AG)20	GCTAGAATGCGAGTCGACAA	59.2	TTGCCTCATCCCTTACGTC	58.7	171
RCM606	27613	(TA)26	ATTTTGAAGCTAAAATGGGCT	57.1	TGTGTAATTGTCTTTTGGGTGTT	58.5	317
RCM609	27613	(TA)20	CGTCTGCTTAAGGCACTGG	59.6	CATTTCTGCTGCTGGGTG	58.9	207
RCM611	27613	(TATG)7	TGGGGACAAAAAGTGGCT	59.1	TGATTCGTGAAGCAGCAGA	59.2	344
RCM613	28637	(AT)49	GGATCCTCAGACGAGGCA	59.9	CAATGCGGTTATTCAATTCG	59.0	254
RCM616	28642	(TA)20	CTAGGCCCCCACCACTCT	60.1	CACGCATATGGCACCAAC	59.5	309

(continued)

Table 5.4 (continued)

Marker	ID	SSR motif	Forward primer (5′–3′)	Tm (°C)	Reverse primer (5′–3′)	Tm (°C)	Expected amplicon size (bp)
RCM623	28644	(TA)45	TGCACCTTTGGAGAACATGA	60.2	TTGGTTTCTGCATTGCCA	59.8	298
RCM634	28657	(ATT)14	TGGAGAATTGTGTCATCTTTCTTG	60.5	CAGCTTGTTTCTTTGCTTTGC	60.2	273
RCM644	28672	(TA)22	AAAAAGAGACGGAATGGCA	58.3	TTGGTACTGTTCGGAGTATGGA	59.5	377
RCM646	28691	(TA)23	GCCATTTTAATAAATTTGTACGTTCTT	59.3	CCTGCCTCCGGTTATCCT	60.0	208
RCM652	28708	(AT)20	AATGCATGGGAGGAAGGA	59.0	CGCTTTAACTAAAGGCCACC	58.9	361
RCM658	27699	(ATA)15	CAAGTTGGCGGACCATTT	59.5	TGCTCACGGGAAGGCTTA	60.5	168
RCM659	27699	(TA)26	TCCAAATTAAAAACTCCAACTCTTTT	59.8	GTCGGAGTCAAAGCGGAA	59.9	384
RCM661	28725	(AT)38	GCTACTGTTGTTGTCTTCGTCG	60.0	GGTAAAAGTTTATTTCATTTGTGTTG	57.5	285
RCM662	28725	(TAT)19	TTCTCCTCAATGCTTGTAAAAAT	57.2	TCGTGTAAGCATTAACATAAATAATTG	58.0	227
RCM668	28732	(ATA)13	GGGGCATACGTGGGTTTT	60.6	GACGTAAATTATCAGGCAAGCTG	60.2	153
RCM670	27715	(TCT)10	GTGTGGCTGTTGGCAGTG	59.9	CCCAATCCAGATGCCCTT	60.8	255
RCM683	28752	(AT)39	TTCCGGGATGGAAATTTTA	57.9	CACAATAAACGAAAATGATTACAACA	59.3	318
RCM684	28752	(ATA)11	TCAATTGCTTGCTTGCCA	60.1	TGGGTGCTATCCCTGCTC	59.7	247
RCM691	27732	(TTA)15	AATTAATCACCGCATTTTTAGC	57.5	CATACAATCACAATTCAATAAAAGC	57.0	371
RCM709	28781	(AG)21	TGCCAAGGCATCCAAGAT	60.2	CAAAGCAATTGCATCCCA	59.2	388
RCM710	27751	(AAT)10	TCACCTTCCCTAAGCTTGAAA	59.0	AACAAACCATCCTTCGTTGTTT	59.8	273
RCM711	27751	(TTA)12	ATGGAGCAATTTGGGCAG	59.6	CCCATCCAGGAATGCTGT	59.4	303
RCM712	29705	(TA)23	ATGATGACCCTGGACCGA	59.8	CAAACCGCTCCTCACCTG	60.4	395
RCM720	27763	(AT)21	CACCATCGGCAATAACCC	59.7	CAACTTTTCCAGATTGAGAACG	58.9	304
RCM725	27770	(TA)30	CATTGAGCTGTTTTCAGGCA	60.0	CGCGTAGGCTGGAAAATG	60.4	207
RCM729	27772	(TAT)13	TCGAATTGGATGGACCTCTT	59.5	CAAACCGATTCGGATCAAA	59.5	160
RCM732	27789	(TTC)13	CGAAATTCAATTGCATCAACA	59.6	ACCGCACGTGACTGACAA	59.9	325
RCM737	27798	(AG)21	GAAGTTCCCCGCCAAAAT	59.9	GGGGCATGATGGACAGAG	60.0	303
RCM741	27798	(AT)29	ATGTGGGTCCTGAGGCAA	60.0	GGGGCTCTAAGTAAATGAGGC	59.2	399
RCM743	27798	(AT)26	GCATGAGGACTATGCCATGA	59.6	TGAAAAATCAGATAAATTGCCA	57.3	374

(continued)

Table 5.4 (continued)

Marker	ID	SSR motif	Forward primer (5′–3′)	Tm (°C)	Reverse primer (5′–3′)	Tm (°C)	Expected amplicon size (bp)
RCM748	29705	(TTA)19	GCGAAAACTCGGCAAAAC	59.4	TCATCCCGAACCACAATATACA	60.1	400
RCM751	29706	(TA)30	TTACCACCTTATAGTCGATCATATCTT	57.9	TGCCGCACTTAATTTTGGT	59.2	243
RCM763	29706	(AT)38	TTTTCAGGCGTGCCTCTT	59.5	CAATGCGGTGTTGAAACTT	57.2	322
RCM767	29706	(AAG)10	GGTAAGGAAGAAAAGTTTGGGG	60.2	CGACGATAGCAGCACCCT	60.0	195
RCM770	29706	(AT)23	GGCAAATGTGGGTCAAGG	59.9	TGCTTTTTGAACTGGCATCA	60.4	357
RCM777	29709	(CT)23	GGATGCTTGCCAGGTTTC	59.2	TCGGCACCTGGGAATTTA	60.0	297
RCM793	29717	(CTT)10	TCAAATGAAGGCCCCAAA	60.0	TTGCATGTTTTTGGAGGGA	60.0	239
RCM802	29724	(TA)22	TGGTTGCATGATTAAATTTAGGA	58.6	GGCTTAGCACTGAGCCGA	60.2	273
RCM812	29724	(TA)33	TCACATTGGAGCCTTATCCC	59.9	CATGTCAATATAAAAATTCTACGTGTT	57.1	227
RCM820	29726	(CTT)12	TCAATTCACGTACGGGGG	60.3	AAAAGCTAGACGTGCCCAAA	59.9	200
RCM821	29726	(TC)23 (TCTA)7[a]	AGTGAAGGGCACAGCACA	58.9	TCAATGCCCCAAAACCTC	59.4	327
RCM822	29726	(AT)26	TGCAGGCCTAACAAAATCAA	59.3	ATCACAACTGCCGCGTTT	60.3	389
RCM824	29726	(AG)21	CAAACCTGCCGAAGTAAAGC	59.9	ATTCAACCGCTGCTACGC	60.0	245
RCM825	29726	(AT)32	TGGGAGACGAAATGAGAAGC	60.3	CGCCATCAAACACCGAA	60.2	387
RCM826	29726	(CT)20	GCCTAAAGGAGCGAAGCC	60.1	TGGGATTGGGAAACCTGA	59.8	329
RCM830	29726	(AG)20	TGGACAAAAAGGGTTGGC	59.5	TTGCAAGGCAATAAGAAAAGAA	59.0	311
RCM850	29729	(TA)32	ATGACCCCGCCCAATAAT	60.0	TTATTGGGTGCCAACAATTT	57.9	266
RCM855	29729	(GAA)18	ACCACCACCATCACCACC	60.1	TGAAGGAGGACACGGGAA	60.2	254
RCM866	29733	(TA)23	GGATTTGGCTTGCCCTTT	60.0	TTTACATGCTCGCACCCC	60.6	303
RCM868	29733	(ATT)15	TTGCGCATATGTAAGCAAAAA	59.4	TTGACAAATAAAATTATGATGGGC	59.2	298
RCM870	29733	(AT)21	CAAAAAGTCATAAAAAGAGTATGATGC	58.8	TTTCTTCCTTGACTCACCTTCC	59.7	356
RCM872	29733	(TA)45	TGGAAATTAGTTTTGAGGGGC	60.3	GGCTCCCATGGTTGAAGA	59.6	343
RCM873	29733	(TAT)13	TTTGTGTAACTTCCCTTCTTCATC	58.7	TTCCACCGAGAAATAAAAGAAAA	59.2	392

(continued)

Table 5.4 (continued)

Marker	ID	SSR motif	Forward primer (5′–3′)	Tm (°C)	Reverse primer (5′–3′)	Tm (°C)	Expected amplicon size (bp)
RCM875	29733	(AT)24	CGGCGTCACTCCTGAAAT	59.8	GGACCCATATAGGGCAAGG	59.2	223
RCM894	29737	(TTA)10	TCATGAGGGCAAATCCGT	60.0	TCTGAAAGGTTGCGGAGC	60.1	215
RCM898	29737	(ATA)10	CGGAAAATAAGAGACATTGCTTG	60.1	TTCTCAAGATGGCATCCACA	60.2	280
RCM900	29737	(AT)27	GATGTGTTTAAAAGTGGCGTTTT	59.5	TGGTCGGACGGTATCCAT	59.7	400
RCM907	29738	(AT)29	GGCAGCTCATATCATTCCCT	59.1	CCTAATGTCCACCGACCG	59.9	230
RCM910	29738	(GAA)13	TATGTTTTCGGGGCAGGA	60.0	CACGCACACGACACAACA	59.9	293
RCM912	29738	(AT)24	AAATGCTGGTTGGGCAAA	60.0	AGAGGGGTAAGGGCATGG	59.9	379
RCM933	29739	(AT)29	AGCCTTTGGATGGAAGCA	59.3	GAAAAAGAAATGTTATTTTAGGAAGAA	57.1	391
RCM934	29739	(AT)27	ATTTTGCCAACCCAACCC	60.6	AAAAGGAAGGTAAGAAACACCAT	57.4	262
RCM935	27810	(TA)22	GGGAATTTCGGTCTGGCT	60.0	CCTGCAATGTCCTGAGCA	59.5	388
RCM938	27810	(AT)22	GGGGAATTGCTGACCCAT	60.7	CCAGACTGGGACACGACC	60.1	367
RCM940	27810	(TA)23	GGAAAACTGTGGCAGCAAA	59.8	TCGCCAGTTTCTACCGCT	59.5	386
RCM941	27810	(TA)21	TGCTGTCTGTGTCATGGAAGA	60.5	GTCCAAACAAAAATTTATATGTCAAAA	58.8	333
RCM943	27827	(AT)22	CAATTGAAAATATAACTCATTAGTCGC	58.7	GGCCTTACGCCTACCCAT	59.9	292
RCM946	27843	(AT)20	GCGGTTATTATTCGCGCT	58.9	GCACACATGTTACAAAAATTAAGCA	60.3	342
RCM948	27849	(AT)27	TTCTCCTCACGCATAGAAATGA	59.8	CAACGGTGGAGTGGGGTA	60.4	372
RCM950	29739	(TA)23	GCCATCGACGAATCAAGC	60.3	AAAAACCACATAAAAGAAAAATTACG	58.3	344
RCM953	29740	(AT)20	GGTGACACCCATGTGCCT	60.4	TTTTTGGGTTGACAGTGGC	59.5	354
RCM954	29740	(TA)22	TCGACTTTTCATTTCATCTTTCC	59.6	TTTCCTAAGCAATAACTTTTTCCA	58.5	344
RCM955	29740	(AT)24	CTCCTACCGGGGCTTGTT	60.1	CACTGATTTGATCGGACTTACG	59.6	360
RCM958	29741	(TA)29	AAAATGTTGTGAAATAAAAACAAACAG	59.0	CATCCAACAGGGAATGTTATGA	59.7	295
RCM963	29742	(AT)30	CCAAAAGATAACGGCCCA	59.5	AATCCTCCTCCTGCCCAC	60.0	264
RCM968	29742	(TA)33	ATCAGTGTTTATCCAAAATAATTGAA	57.4	CAGAGGTACAATGCAGAAATGAA	59.3	284
RCM994	29747	(TA)26	CAGGTATATAATTCTCTGATGCTTGC	59.6	CCTCATATTTTCTTTCCAATATTCA	57.8	171

(continued)

Table 5.4 (continued)

Marker	ID	SSR motif	Forward primer (5′–3′)	Tm (°C)	Reverse primer (5′–3′)	Tm (°C)	Expected amplicon size (bp)
RCM1000	29751	(AT)26	TCATTATTAAAGAAAAATTTCGGG	57.8	ACGGAAAGCATATCGATTTTT	57.8	179
RCM1001	29751	(AAG)13	TGGCAGAAGGAGAATGGC	59.9	GGCAAAGTTACTGGTGACCC	59.5	291
RCM1003	29751	(AT)22	AACGGATAGGGTACCCACCTA	59.6	TGGGTTCGGATAATATACGGG	60.7	190
RCM1009	29751	(TA)25	AATGGGCCAACACCTTTG	59.3	GAATCTCGTTTTCCGGGG	60.4	241
RCM1010	29751	(AT)24	TCGGACACAGGAACTCCG	60.8	TGGCAGGGACAACAACAA	59.6	380
RCM1022	29757	(AT)26	CTTGCCATCTGGCTCAAA	58.4	TGCCATTGGTGGTTCTCC	60.5	324
RCM1037	29761	(AT)20	TGGAAATTTTATTTCTCCAAGCA	60.0	TCAGTTTGTGCATTAAGAATTGG	59.2	334
RCM1040	29762	(TA)20	TGAAAGAGCTTCACTGCCC	59.1	TGGTTCATCCTGGCCATA	58.3	389
RCM1042	29763	(AT)20	ATCGATCAGCCGAACCAA	60.2	CGTGTGAGAAGCTGCGGT	61.2	180
RCM1047	29763	(AAAG)7	TTGGTGGCTTTTTGAGCTG	60.0	TGCAATGATTTTGGGTGC	59.0	254
RCM1053	29768	(TA)27	TCCAGGATCCCTTTGCAC	59.6	CATCCCGTGAGACTTGGC	60.2	302
RCM1054	29768	(TTA)13	GAGAGAGTGAAAGTGTCAAACACC	59.3	GGCTTATTGGGCACAGGA	59.6	215
RCM1056	29771	(ATT)13	AGCGGCAACAAGCTATGG	60.0	CATTGGTGCTGCTGGAAA	59.3	400
RCM1060	29773	(AT)23	TCTCATCTGTCAAAAATAAATACACA	57.1	TTGCCGAGTCGTATGAGG	58.3	218
RCM1065	29776	(AT)21	AGTGCGCATGAGGTAGTCG	60.0	GAATCTTGTGAACCGCATGA	59.7	376
RCM1074	29780	(AT)26	TTGACTTGCTTTTAGCTAAGCCT	59.0	TCAATCGAAGACGGAGCC	59.9	270
RCM1075	29780	(GAA)13	TGATCGAGGGAGGCACAT	60.2	CAACGGGTCCATCCCTTA	59.3	242
RCM1076	29780	(TG)23	CGAATGCTGTCCCCAAAC	60.0	TGTCATGGATGCACACACA	59.0	364
RCM1080	29780	(AT)32	GAGGAACAAACAATGGGGTTT	60.1	ATGGGACGTGTGAGTGCC	60.6	374
RCM1088	29785	(ATA)12	TTGTCCCAGCCACCACTT	60.1	TTTCATGCCAAAAATCACCA	59.9	231
RCM1093	29785	(TTA)12	TGGTGGTTGGACACTCTTTTC	60.0	GACTTTGATAATACATCTTGGGTGC	60.1	258
RCM1095	29785	(AAT)13	CATTCACTCTTGGCCAGTCA	59.8	ATTATGGTGCGCGGATTG	60.5	185
RCM1311	29785	(ATA)12	GGCATTCACCACCTTCTTGT	60.0	TTTCATGCCAAAAATCACCA	59.9	206
RCM1313	29785	(TAT)14	CCACAACCATCACACCTCAG	60.0	CCGGTTGATTTTGTGTACGA	59.4	288

(continued)

Table 5.4 (continued)

Marker	ID	SSR motif	Forward primer (5′–3′)	Tm (°C)	Reverse primer (5′–3′)	Tm (°C)	Expected amplicon size (bp)
RCM1316	29785	(TAT)… (TTA)10	TTGAGGTGAGTAGTGAGCAATTTT	59.4	CGCGACAAATAATTCAGTGG	59.2	399
RCM1317	29785	(TTA)12	TGGTGGTTGGACACTCTTTTC	60.0	GACTTTGATAATACATCTTGGGTGC	60.1	258
RCM1319	29785	(AAT)13	CATTCACTCTTGGCCAGTCA	59.8	CGGATTGCAAGGATAAGCC	60.6	174
RCM1335	29792	(GA)34	ATTTGACGACGAAAACGGTC	60.0	GATGGGAATGCGATCAATCT	59.9	328
RCM1336	29792	(TG)24	GGCCCAAATCTGAGCAGTAA	60.2	CACACTCACCCAGCACACTC	60.4	386
RCM1338	29792	(AG)29	AGCTTCAGGTCCACCATCAC	60.1	TTCAAGAGGGCTAGCAAGGA	60.1	374
RCM1341	29794	(TTTTA)7	GGGGTTTTTCGGTTTTGATT	60.0	CAGATGAAAATCTGATATGGCAA	59.1	273
RCM1343	29794	(AT)29	AATGGCAAATGAGTGCCTTC	60.1	CGAATGACCAAATGCTTCAA	59.7	207
RCM1350	29794	(TA)21	TGCAGAAGCATAAGTTTCCAA	58.6	CTATGTGCCATAGGCCAGGT	60.0	396
RCM1351	29794	(GA)22	ATGTTTCCGAGCAAAGGAGA	59.8	ATCCAGCTCGCCATTATCAC	60.1	376
RCM1382	29801	(ATT)14	ATGTTTGCAAATGTCAGGCA	60.1	TAAGTCCGGGGTGAGAAAAA	59.5	385
RCM1386	29801	(TTA)10	GCAGTGCAATCATGGTTAATG	59.1	ATGGGTTGCTGCTGGTTTTA	60.5	284
RCM1394	29801	(TA)22	AGATGACAGCCATCACCCAT	60.4	AAGCGTTACGTGCAAGTGTG	60.0	281
RCM1395	29801	(AT)24	GGAAAATTGCAAGAATTTGACA	59.1	GGGGGTTTGATGTTACCCTT	59.9	389
RCM1396	29801	(AT)24	TAGTGGGACCACTGACCCAT	60.2	GTTTTGAGCCAATCAATGCC	60.5	360
RCM1413	29804	(TA)31	TGGGTATTGGGTACCCTCTTT	59.6	TGCTTCATATGTGGCATGTTT	59.1	379
RCM1424	29806	(GA)21	GCATCAAGCAGAGTGCCATA	60.0	TATTCCAAGAGCATCCCTGC	60.2	383
RCM1434	29807	(TA)28	GGTTGGGTCCAGTAGCAAAA	60.0	CTACGTGAAACCCCTGCATT	60.0	383
RCM1444	29808	(AT)34	ATGAACCACGCTCTCAGCTT	60.0	TCTTGTGAAAAGCATCGCAT	59.4	368
RCM1450	29811	(TA)26	TCCCGAATTTATACTTCTATTCAAGG	60.1	TTTGAAATTCATACCCATATCGAG	59.3	382
RCM1463	29813	(AG)24	GGCACTTAGGACCACTTGCT	59.4	ATAAAACATGGTTCAGCGGC	60.0	270
RCM1466	29813	(AT)24	TCGGGAATGAACTGTTCGAG	61.2	AGGATTTTGAAGCAATGGCA	60.6	363

(continued)

Table 5.4 (continued)

Marker	ID	SSR motif	Forward primer (5′–3′)	Tm (°C)	Reverse primer (5′–3′)	Tm (°C)	Expected amplicon size (bp)
RCM1467	29813	(TA)27	AAAACTTTCTGCAAACAACAATTT	58.4	TCAAAAGATTTCATAACTCGTGATT	58.3	388
RCM1468	29813	(AT)27	TTGAAGCTCCCAAAACAAGAA	59.8	TCAATTTGCAAGGTGTAGCC	58.8	318
RCM1471	29813	(TA)23	GGCATATTCATGCTGCGATA	59.7	TAAACAATGTCCGAGAGGGG	59.9	292
RCM1472	29813	(GAA)10	CAAGCACTTTAACCGCACAA	59.9	AGCCCTTCTCTTTCCTTTGC	60.0	210
RCM1473	29814	(AGA)10	GAGGAAGACCACTGCCATGT	60.1	CAAGTGGGTCCTTGAAGATGA	60.1	291
RCM1481	29815	(CCA)12	GCGACACCTAATAGCTGCAA	59.1	TGACCCAAGAGAAAGAGGGA	59.8	356
RCM1492	29816	(AT)44	TTTAATAAAATTGGTGGTCTTTTGAA	59.3	TCAAAGAATTTGGAAGCTAAAGAAA	59.8	327
RCM1493	29816	(TA)20	TTTCATATCCCACTCCCACC	59.6	ACGCATTGGACCGTTATTTT	59.3	283
RCM1495	29816	(GA)30	TTCAACCGCACAATTGCTAA	60.3	AGCAAGAGTGGTGAGACGGT	59.9	377
RCM1496	29816	(TA)22	GTGGAGGCAAACACAATCCT	60.0	AGTATATGGCCCCCACTCAA	59.3	262
RCM1497	29817	(AGA)13	AGCCACACATGGTGGATTAAG	59.9	ACCCATGAGCTGCTGAAACT	59.9	373
RCM1500	29818	(TA)40	CAGGACATGTGTCAATCATCAA	59.4	TTGGACATTTTCATTTTGGGA	60.2	300
RCM1503	29820	(TTC)10	TGGAGGCTTGCCCTAATGTA	60.6	CGTGCTACAATTCTTCTGCG	59.6	134
RCM1510	29822	(TAA)14	TTGGAAATTGAGACTATTGACCC	59.4	TTCATTTGGCAGAAGATTTTGA	59.7	299
RCM1516	29822	(AT)26	GGTGGTGCTGTTCACGTATG	60.0	GTAGAAGGCAACCAAGTGCG	60.8	348
RCM1519	29822	(TA)31	TGAATGTTGTAATGGAGCCG	59.5	CAGCCATTTGGTTATTTGGAA	59.8	387
RCM1520	29822	(AT)31	AATTAATGGAATTGGACGCC	58.8	CGGACGGATCTTGTGTCATA	59.5	353
RCM1521	29822	(AAT)11	CTAAGGAGTTGTCTTCGCCG	60.0	GGCAAAGGTTTGACAGTGGT	60.0	358
RCM1522	29822	(AGA)10	GCGTTTCCTTCCCTTTTACC	59.9	AACCATGCATGCAAAATCCT	60.3	390
RCM1524	29822	(AT)21	TGATGGTAAATTGCTTGGCA	60.1	TGCAGGGAACAGGTAGTTGA	59.3	161
RCM1526	29822	(TAT)21	AATAGCATGATGGTCGGTCC	59.8	TGTCCCCTTTTCCATTCTTG	59.9	335
RCM1527	29822	(TA)24	AAGCTTGAATCTCATGTGTCAAA	58.9	TGGGTCTTCAGGATCGTTTC	60.1	330
RCM1534	29823	(TAT)18	TCATCCCATTTAAAATAAAAACTTTG	58.9	GAAGCATGAGAGTTTTTCTCCC	59.4	277

(continued)

Table 5.4 (continued)

Marker	ID	SSR motif	Forward primer (5′–3′)	Tm (°C)	Reverse primer (5′–3′)	Tm (°C)	Expected amplicon size (bp)
RCM1540	29825	(AT)22	CCCTATGGAGCTTTGACCCT	60.5	GTTGGTCGACTTCGAGTGGT	60.2	390
RCM1545	29826	(AT)28	GGACCATTACACGCAACTGA	59.6	CGGAGTGAAGAAGCCATCTC	60.0	397
RCM1547	29827	(AT)21	GCTGATAGCTGATTTAAATTTATTTGT	57.8	GGTGGCTGATAGCCGATAAA	60.1	338
RCM1552	29827	(GAA)10	ACTTTTGCTTCTTCCCACGA	59.9	GTGCCTGTAACCGAGATGGT	60.0	225
RCM1553	29827	(AT)25	TTTATGCCCAGAAGACCTGG	60.1	GCGATGCAAGGGTCAAGTAT	60.1	272
RCM1555	29827	(TC)25	TGCACATTAGCCTGAAGGAA	59.4	ATTGCTCCTCCAACACCATC	59.9	347
RCM1567	29828	(AT)22	CTGCCTCTGTAGGGTTCTGC	60.0	CTAGGTAGTGAATGCCCCCA	59.9	346
RCM1581	29835	(TC)23	ATCACAAGGCAACCCAAGTC	60.0	GCAGTTCCATTCCTCTTTGC	59.8	185
RCM1582	29835	(AT)26	GCACGTAAAAAGCCCAAAAA	60.1	GGGTTTATGACGGTGCATTC	60.2	318
RCM1588	29835	(TA)31	TTATGCTATTGTTTTTAGGGGG	57.3	TGGTGCAGAGGTGGCTTAGT	60.9	395
RCM1594	29836	(TCT)13	TGATTGTCTCGGTCTTGCAG	60.0	CCGAGGTTGATCACCAAGAT	59.9	332
RCM1600	29836	(GA)29	ATCTCTGCATCTGAGCGTGA	59.7	ATCCAGCGGATGCTACTGAG	60.4	369
RCM1601	29836	(AT)31	AGCTAGATGGTGCAAGGTCAA	59.9	CTTACAAAGCCCCTCCATGA	60.1	336
RCM1602	29836	(ATA)23	CCTTGACCAGTTGGTTCTCC	59.5	CCAATATGGGCTCCAGTCAT	59.8	171
RCM1603	29836	(AT)23	TCAAAACCAAACTCCAACCC	59.8	AAACGACGCCGCAATTATAC	60.0	313
RCM1608	29838	(ATA)14	GCAAAGCAGGAAAGGTCACT	59.5	TGCAATTAGCTTGGCAATGA	60.4	398
RCM1625	29839	(TAT)13… (TTA)12	CGAGGATTGATCCGAACTGT	60.1	AGCACGTTTTACTGCCATCC	60.1	378
RCM1626	29839	(ATT)15	GCCGGTTTATTAATTTTGAAAG	57.1	ATGACGTCTTCCTATGCGCT	59.9	379
RCM1638	29840	(TAT)13… (TAT)12	GCATGAGCAACGCATAAGAG	59.6	CATTTATGGAAGAACAACAACCA	58.9	400
RCM1641	29841	(TA)23	CTGGGCTAGAGCATTGGAAG	60.0	TCAATTTGTGCATTAGGAATTTG	59.0	380
RCM1642	29841	(TAT)10	CCCCGATAACCAATTAAACA	57.4	GCACACAAAAAGACAAATGAAGA	59.3	400
RCM1643	29841	(AT)23	TGCATGGCTTGAGCTATCTG	60.1	AAGGCTGACATGCACCACTA	59.3	300

(continued)

Table 5.4 (continued)

Marker	ID	SSR motif	Forward primer (5′–3′)	Tm (°C)	Reverse primer (5′–3′)	Tm (°C)	Expected amplicon size (bp)
RCM1648	29841	(CT)20	ATGGAGCAACTTGTCGCTTT	59.9	AAACAACACATCCACATCGC	59.4	360
RCM1649	29841	(TA)23	CAAAAATCAAGCACACATGTCA	59.6	CGTGAAATATGTGAAACGATGAA	59.9	370
RCM1651	29841	(TA)28	GTAGCCAAAGTGGGCAAAAA	60.1	GAGATTTCCCCAAACCACCT	60.2	315
RCM1698	29844	(TTA)13	AAACCTCCATGGTGCACTTC	60.0	GTTCGATTTGTTGGCGAGAT	60.1	294
RCM1699	29844	(TA)20	TAATTGGTGTGGGCTGTGAA	60.0	AACACCGAGATTGAGGATGG	59.9	232
RCM1700	29844	(TA)28	TGCCAACTTTCAAGACCAAA	59.3	AAATTTTGTGCATGTTGGGA	58.9	369
RCM1702	29844	(AT)34	TTGATCCAATATATGCAGGATTT	57.6	CCACAAAAAGGCATTTGTCA	59.6	241
RCM1704	29844	(CAA)19	CTTGCAAAGTCAGGCAAACA	60.0	GTTGGTTCTGGTTTGGTTGG	60.2	232
RCM1705	29844	(AT)23	TTGTGTGGTCAGCACTCCTT	59.3	TTTTTGAAAACTGCACAGCG	60.0	320
RCM1708	29844	(AT)24	AGCAAGGTGGCACCAGTATC	60.1	CCTTACGTCCCTACCTTCCA	59.0	283
RCM1709	29844	(TA)39	GTGCGTTATGCTGGTTTGTG	60.2	GAGACGGTGCACATGGATAA	59.5	263
RCM1710	29844	(AT)37	ACGTGCATGTAAGATTTTTGC	57.8	AAGTTTTGTTGCACTCACCG	58.8	286
RCM1713	29844	(ATGT)7	TTGTTCGCTCTTTGGGAAGT	59.9	AGCAAGCCTCCACCAACTAA	59.9	389
RCM1716	29846	(TA)35	TACACGCGTGCAACTTCTCT	59.7	GCTTGTAAGACAGGCTCCGA	60.5	367
RCM1717	29846	(GA)23	TTTGAGCATGCAAAGATTCA	58.0	TTGCATGTTGCATTCAGTCA	59.8	299
RCM1727	29848	(TAT)13	TGATCAATCGAAATGAACCG	59.5	GGTCGGCTAGGTTTGACAGA	60.3	173
RCM1728	29848	(AAACA)8	GAGCTGTGGCTGACAATGAA	60.0	TCTCCTAGGCCAATCCTTGA	59.8	157
RCM1731	29848	(TTA)18	AACTTTATGCAAGTGGGAAACAA	59.9	CCCTCATTTCTGGCCTATGA	60.0	299
RCM1733	29848	(CT)21(AT)24	ATTTCGTGGCTGCCTATGAT	59.6	AGATGGCGTCATGTATAGTTGG	59.0	294
RCM1738	29848	(AT)29	CAAGATTCGGAGGAGGATCA	60.2	AGAATGGACGACGCAGAACT	59.9	305
RCM1739	29848	(CTT)14	AAATACACCTTGCGCTCTCG	60.4	TGCCATTTATCACTGCCTTG	59.7	235
RCM1767	29852	(TA)24	GCCAATGGGCTCTACTTTGA	60.2	GTCAGGAAAAGGGGTGTTGA	59.9	321
RCM1768	29852	(AT)27	ATTTTCTAATGCCGGTCACG	60.0	GCGTCTGTTGTCTTGTTGGA	59.9	385

(continued)

Table 5.4 (continued)

Marker	ID	SSR motif	Forward primer (5′–3′)	Tm (°C)	Reverse primer (5′–3′)	Tm (°C)	Expected amplicon size (bp)
RCM1769	29852	(TA)20	CGAAAACAAAAGAATGAATTGC	58.8	ACCAATACGTCCATTCCTGC	59.8	385
RCM6085	27798	(TA)28	CGATTCAATCAGGACCCTA	58.4	AGTAAATGAGGCTCCACATA	56.7	325
RCM6086	27798	(GA)20	GAATAGAAGTCTAACCACAGAG	59.4	GCTTCTTGCGGCTATACTA	58.4	143
RCM1201	[a]	(AG)20	TGAACAGTTTAAATGAACTCCA	50.0	GAACCATATCTTGCCTGTTTAT	50.3	176
RCM1203	[a]	(CT)12	GCGGTGTTATCATTCCTATATT	50.0	AATCCTCATTTTGCATTCTTAC	49.2	301
RCM1204	[a]	(GTT)9	CACTTGGATCGAACTCAGAC	51.7	TGATTTGCTCTTTGTTTATCAG	49.2	272
RCM1210	[a]	(TTC)4, (TTC)5	CATATGTGATCAATAGGGATGT	49.3	AATCCAGTAGATTTCTCATTGC	50.4	265
RCM1212	[a]	(GCA)4, (GAA)7	AAATGGAAGATCATAAAACTCG	48.8	CTCCTTCTCTTGGAAATTACTG	50.4	275
RCM1220	[a]	(TTC)7	ATCTTTTCTTGAACTGTGACCT	51.7	CATGGTGTAGATTCATCATTGT	50.4	200
RCM1222	[a]	(CAG)5	TCCAACAAGGATTAGACTTTCT	50.8	AAGTTGTCTGTTGTTGTAGCTG	53.3	290
RCM1223	[a]	(AG)11	TGTTTAGAAGCAACTCATCTTG	50.9	TTCAGGGTGTTTTTAGACTTTT	50.1	302
RCM1227	[a]	(GAA)11	AAAAAGTAGCAGAACTAGCACC	52.8	ACCATCTTCTTCATCTTGTTTC	50.5	185

[a]RCM1201, 1203, 1204, 1210, 1212, 1220, 1222, 1223 and 1227 were cited from Seo et al.'s study (2011), which original names were GB-RC-025, 046, 059, 225, 001, 107, 133, 135 and 217, respectively

Table 5.5 Molecular markers used for construction of genetic linkage map

Types of markers	Number of primers screened	Number of polymorphic primers	Number of polymorphic markers	Polymorphism (%)	Number of mapped markers
RAPD	520	119 (22.80%)	127 (1.06 marker/primer)	24.42	76 (59.84%)
ISSR	100	43 (43.00%)	47 (1.09 marker/primer)	47.00	34 (72.34%)
SSR	300	162 (54.00%)	162 (1 marker/primer)	54.00	151 (93.20%)
Total	920	324 (35.21%)	336 (1.03 marker/primer)	37.60	261 (77.68%)

Table 5.6 Salient features of genetic map of castor construction of genetic linkage map

LG	No. of mapped markers				Density (Marker/cM)	No. of interval (>1 cM)	No. of gaps (>10 cM)	Average marker interval (cM)	Length (cM)
	PAPD	ISSR	SSR	Total					
LG1	4	3	15	22	0.13	21	5	7.34	161.5
LG2	3	5	14	22	0.15	21	1	6.70	147.5
LG3	5	4	15	24	0.12	23	8	8.23	197.5
LG4	11	8	16	35	0.15	34	3	6.64	232.5
LG5	12	2	13	27	0.14	26	7	7.11	192.2
LG6	16	0	16	32	0.14	31	5	7.30	233.7
LG7	7	3	14	24	0.17	23	1	5.82	139.8
LG8	8	1	10	19	0.15	18	3	6.76	128.5
LG9	5	5	24	34	0.12	33	16	8.19	278.7
LG10	5	2	14	21	0.17	20	0	5.78	121.5
Total	76	34	151	261	–	250	49	–	1833.4
Average	7.6	3.4	15.1	26.1	0.14	–	5.7	6.93	–

The length of the 10 LGs ranged from 117.88 cM (LG5) to 190.47 cM (LG3), with an average of 154.74 cM. The largest LG was LG4 with a genetic length of 148.50 cM. LG4 contained 1064 markers and had an average marker density of 0.14 cM. The smallest LG was LG5, with a genetic length of 117.88 cM, 149 markers, and a 0.80 cM marker interval.

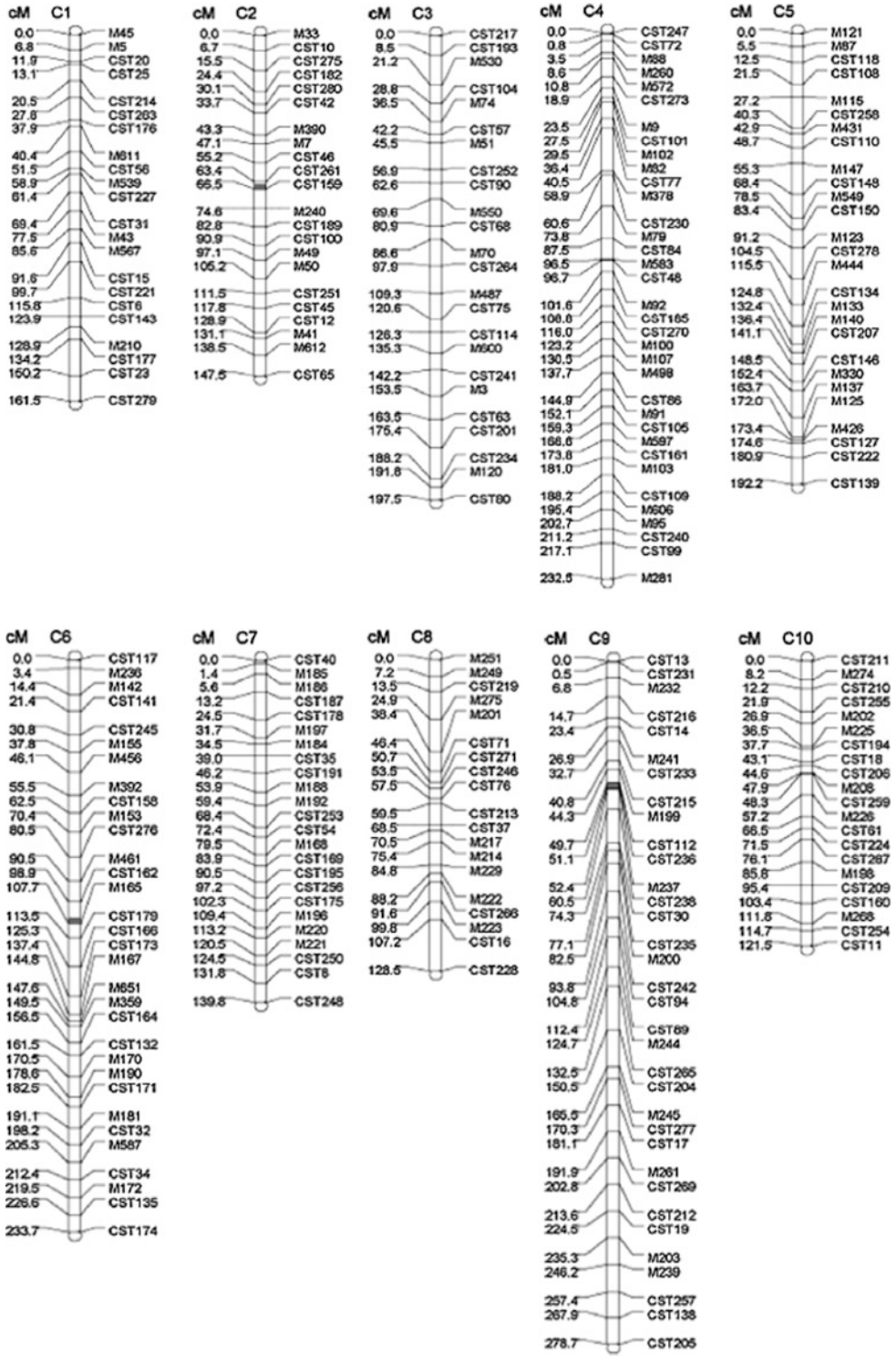

Fig. 5.2 Linkage map of *Ricinus communis* L. The C refers to the chromosome corresponding to the LG with the same code

Table 5.7 Sequence coverage of SLAF markers

Samples	SLAF number	Total depth	Average coverage
YC2	117,025	2,543,514	21.73
YF2	146,843	4,546,512	30.96
Offspring	108,686	697,910	6.42

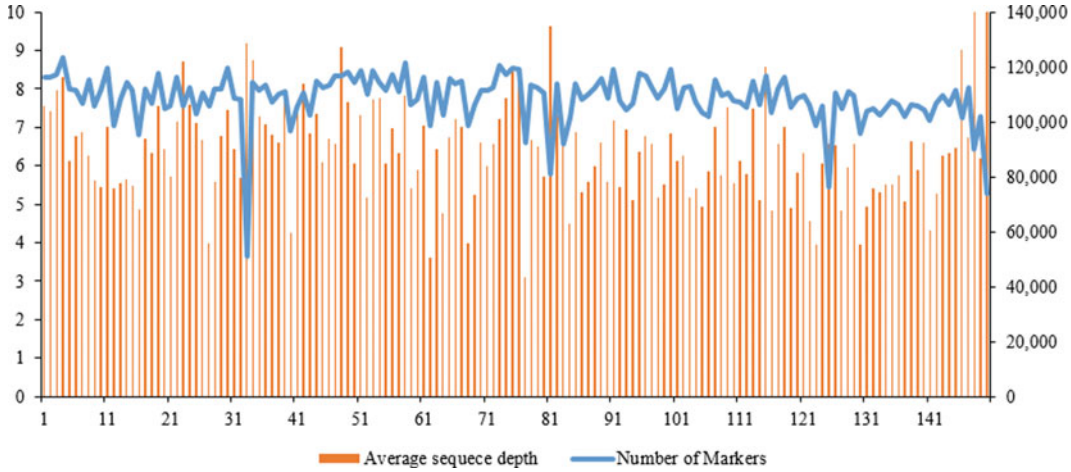

Fig. 5.3 Number of markers and average sequencing depths in the F_2 population. The *x*-axis indicates the individuals, the left, and right *y*-axes indicate average sequence depth and number of markers, respectively

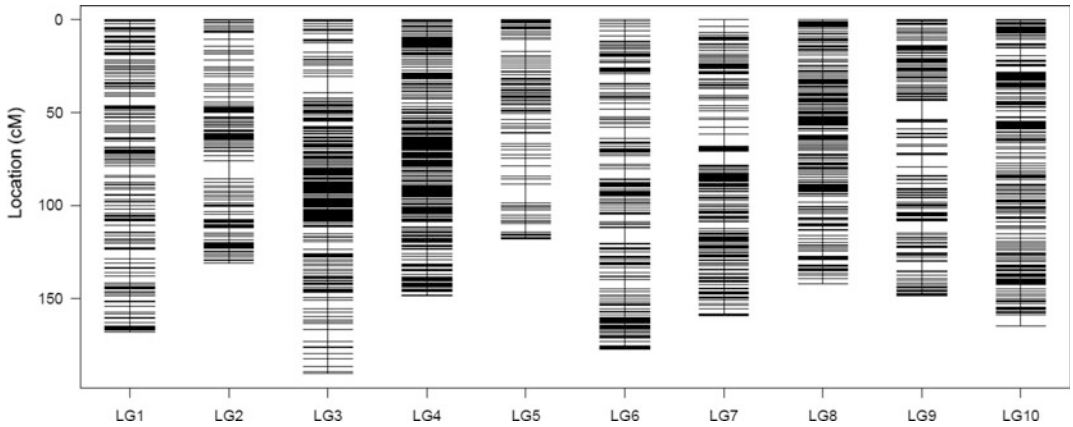

Fig. 5.4 High-density genetic map of castor bean

The largest gap on this map was 10.13 cM located in LG9 (Table 5.8). To our knowledge, this is the densest genetic map of castor bean.

The SLAF-based castor bean genetic map was used to anchor the sequence scaffolds that compose the draft sequence of the castor bean

Table 5.8 Characteristics of the ten LGs in the SLAF-based genetic map

Linkage group ID	Total marker	No. of SSR marker	Total distance (cM)	Average distance (cM)	Max gap	SNP number	No. of anchored scaffolds	Scaffold length
LG1	272	11	167.79	0.62	5.35	344	55	22,278,757
LG2	179	13	131.02	0.74	9.74	230	32	13,871,998
LG3	707	17	190.47	0.27	8.86	1028	87	20,944,873
LG4	1064	16	148.5	0.14	2.54	1436	97	26,275,325
LG5	149	0	117.88	0.80	10.12	184	23	19,792,242
LG6	346	27	177.25	0.51	8.45	420	53	23,652,170
LG7	387	13	159.06	0.41	7.39	486	43	21,427,524
LG8	684	9	142.08	0.21	3.2	923	76	23,966,805
LG9	207	12	148.54	0.72	10.13	252	43	14,362,120
LG10	425	2	164.82	0.39	6.23	555	57	28,132,081
Total	4420	120	1547.41	0.35	10.13	5858	566	214,703,895

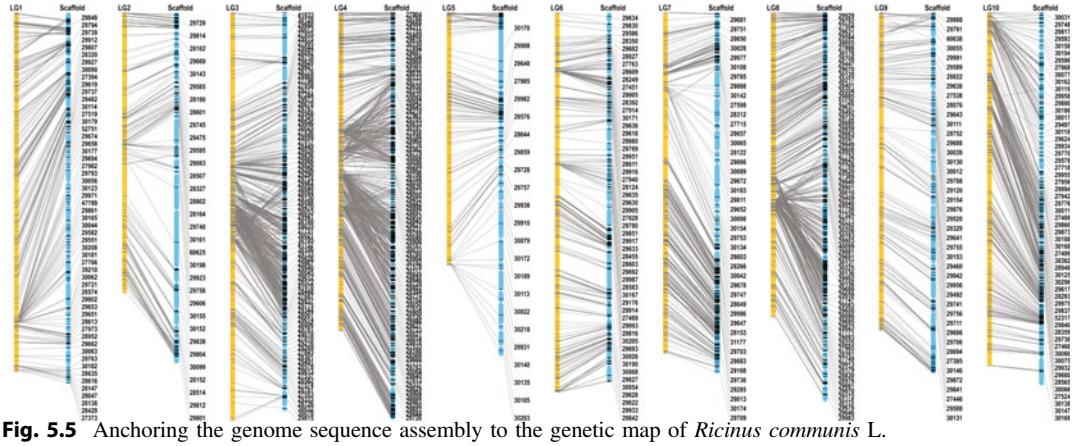

Fig. 5.5 Anchoring the genome sequence assembly to the genetic map of *Ricinus communis* L.

genome (Chan et al. 2010) and 566 (2.19%) of the scaffolds, covering 61.14% of the genome, could be linked to markers in the genetic map (Fig. 5.5). The average ratio of genetic-to-physical distance was 1 cM per 226 Kb, which is very similar to that observed in *Jatropha curcas* L. (1 cM per 251 Kb) (Wu et al. 2015). The scaffolds were assembled into ten pseudochromosomes, which represents a useful resource for comparative genomics analyses in the Euphorbiae family.

References

Allan G, Williams A, Rabinowicz PD, Chan AP, Ravel J, Keim P (2008) Worldwide genotyping of castor bean germplasm (*Ricinus communis* L.) using AFLPs and SSRs. Genet Resour Crop Evol 55:365–378

Bajay MM, Pinheiro JB, Batista CEA, Nobrega MBM, Zucchi MI (2009) Development and characterization of microsatellite markers for castor (*Ricinus communis* L.), an important oleaginous species for biodiesel production. Conserv Genet Resour 1:237–239

Bajay MM, Zucchi MI, KIIhl TAM, Batista CEA, Monteiro M, Pinheiro JB (2011) Development of a

novel set of microsatellite markers for Castor bean, *Ricinus communis* (Euphorbiaceae). Amer J Bot 98: e87–e89

Bennetzen JL, Freeling M (1997) The unified grass genome: synergy in synteny. Genome Res 7:301–306

Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, Melake-Berhan A, Jones KM, Redman J, Chen G, Cahoon EB, Gedil M, Stanke M, Haas BJ, Wortman JR, Fraser-Liggett CM, Ravel J, Rabinowicz PD (2010) Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotechnol 28:951–956

Chen C, Zhou P, Choi YA, Huang S, Gmitter FG Jr (2006) Mining and characterizing microsatellites from citrus ESTs. Theor Appl Genet 112:1248–1257

Cordeiro GM, Christopher MJ, Henry RJ, Reinke RF (2002) Identification of microsatellite markers for fragrance in rice by analysis of the rice genome sequence. Mol Breed 9:245–250

Devos KM, Gale MD (1997) Comparative genetics in the grasses. Plant Mol Biol 35:3–15

Diaz A, Fergany M, Formisano G, Ziarsolo P, Blanca J, Fei Z, Staub JE, Zalapa JE, Cuevas HE, Dace G, Oliver M, Boissot N, Dogimont C, Pitrat M, Hofstede R, van Koert P, Harel-Beja R, Tzuri G, Portnoy V, Cohen S, Schaffer A, Katzir N, Xu Y, Zhang H, Fukino N, Matsumoto S, Garcia-Mas J, Monforte AJ (2011) A consensus linkage map for molecular markers and quantitative trait loci associated with economically important traits in melon (*Cucumis melo* L.). BMC Plant Biol 11:111

Feuillet C, Keller B (2002) Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution. Ann Bot 89:3–10

Foster JT, Allan GJ, Chan AP, Rabinowicz PD, Ravel J, Jackson PJ, Keim P (2010) Single nucleotide polymorphisms for assessing genetic diversity in castor bean (*Ricinus communis*). BMC Plant Biol 10:13

Goldstein DB, Schlotterer C (1999) Microsatellites: evolution and applications. Oxford University Press, Oxford

Gupta K, Balyan S, Edwards J, Isaac P, Korzun V, Roder M, Gautier MF, Joudrier P, Schlatter R, Dubcovsky J, De La Pena C, Khairallah M, Penner G, Hayden J, Sharp P, Keller B, Wang C, Hardouin P, Jack P, Leroy P (2002) Genetic mapping of 66 new microsatellite (SSR) loci in bread wheat. Theor Appl Genet 105:413–422

Helentjaris T, Weber D, Wright S (1988) Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms. Genetics 118:353–363

Liu S, Yin XG, Lu JN, Liu C, Bi C, Zhu HB, Shi YZ, Zhang D, Wen DY, Zheng J, CuiY Li Wj (2016) The first genetic linkage map of *Ricinus communis* L. based on genome-SSR markers. Indust Crops Prod 89:103–108

Machado EL, Silva SA (2013) Design and validation of SSR microsatellite primers for castor bean. Pesqui Agropecu Bras 48:1457–1463

McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 9:199–207

Milbourne D, Meyer R, Collins A, Ramsay L, Gebhardt C, Waugh R (1998) Isolation, characterisation and mapping of simple sequence repeat loci in potato. Mol Gen Genet 259:233–245

Moore G, Devos KM, Wang Z, Gale MD (1995) Cereal genome evolution—grasses. Line up and form a circle. Curr Biol 5:737–739

Morgante M, Salamini F (2003) From plant genomics to breeding practice. Curr Opin Biotechnol 14:214–219

Pranavi B, Sitaram G, Yamini K, Dinesh Kumar V, Van Deynze A (2011) Development of EST-SSR markers in castor bean (*Ricinus communis* L.) and their utilization for genetic purity testing of hybrids. Genome 54:684–691

Qiu L, Yang C, Tian B, Yang J-B, Liu A (2010) Exploiting EST databases for the development and characterization of EST-SSR markers in castor bean (*Ricinus communis* L.). BMC Plant Biol 10:278

Rallo P, Dorado G, Martin A (2000) Development of simple sequence repeats (SSRs) in olive tree (*Olea europaea* L.). Theor Appl Genet 101:984–989

Seo KI, Lee GA, Ma KH, Hyun DY, Park YJ, Jung JW, Lee SY, Gwag JG, Kim CK, Lee M (2011) Isolation and characterization of 28 polymorphic SSR loci from castor bean (*Ricinus communis* L.). J Crop Sci Biotechnol 14:97–103

Sharopova N, McMullen MD, Schultz L, Schroeder S, Sanchez-Villeda H, Gardiner J, Bergstrom D, Houchins K, Melia-Hancock S, Musket T, Duru N, Polacco M, Edwards K, Ruff T, Register JC, Brouwer C, Thompson R, Velasco R, Chin E, Lee M, Woodman-Clikeman W, Long MJ, Liscum E, Cone K, Davis G, Coe EH Jr (2002) Development and mapping of SSR markers for maize. Plant Mol Biol 48:463–481

Song Q, Shi J, Singh S, Fickus E, Costa J, Lewis J, Gill B, Ward R, Cregan P (2005) Development and mapping of microsatellite (SSR) markers in wheat. Theor Appl Genet 110:550–560

Song Q, Jia G, Zhu Y, Grant D, Nelson RT, Hwang EY, Hyten DL, Cregan PB (2010) Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1. 0) in soybean. Crop Sci 50:1950–1960

Sun X, Liu D, Zhang X, Li W, Liu H, Hong W, Jiang C, Guan N, Ma C, Zeng H (2013) SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. PLoS ONE 8:e58700

Tan ML, Wu K, Wang L, Yan MF, Zhao ZD, Xu J, Zeng Y, Zhang XK, Fu CL, Xue JF, Wang LJ, Yan XC (2014) Developing and characterising *Ricinus communis* SSR markers by data mining of whole-genome sequences. Mol Breed 34:893–904

Tomar RS, Parakhia MV, Rathod VM, Thakkar JR, Padhiyar SM, Thummar VD, Dalal H, Kothari VV, Kheni Jasminkumar, Dhingani RM, Sabara Pritesh, Golakiya BA (2017) Molecular mapping and identification of QTLs responsible for charcoalrot resistance in Castor (*Ricinus communis* L.). Indust Crops Prod 95:184–190

Tuberosa R, Salvi S (2006) Genomics-based approaches to improve drought tolerance of crops. Trends Plain Sci 11:405–412

Van Ooijen J (2006) JoinMap 4. Software for the calculation of genetic linkage. Maps in experimental populations. Plant Research International

Wei F, Coe E, Nelson W, Bharti AK, Engler F, Butler E, Kim H, Goicoechea JL, Chen M, Lee S, Sanchcz-Villeda H, Schroeder S, Fang Z, McMullen M, Davis G, Bowers JF, Paterson AH, Schaeffer M, Gardiner J, Cone K, Messing J, Soderlund C, Wing RA (2007) Physical and genetic structure of the maize genome reflects its complex evolutionary history. PLoS Genet 3:1254–1263

Wei X, Wang L, Zhang Y, Qi X, Wang X, Ding X, Zhang J, Zhang X (2014) Development of simple sequence repeat (SSR) markers of sesame (*Sesamum indicum*) from a genome survey. Molecules 19:5150–5162

Wu YH, Close TJ, Lonardi S (2011) Accurate construction of consensus genetic maps via integer linear programming. IEEE/ACM Trans Comput Biol Bioinform 8(2):381–394

Wu P, Zhou C, Cheng S, Wu Z, Lu W, Han J, Chen Y, Chen Y, Ni P, Wang Y (2015) Integrated genome sequence and linkage map of physic nut (*Jatropha curcas* L.), a biodiesel plant. Plant J Cell Mol Biol 81:810–821

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. Mol Ecol 11:1–16

Genomics of Disease Resistance in Castor Bean

6

Archit Sood and Rajinder Singh Chauhan

Abstract

Castor bean is considered an economically important plant species within the Euphorbiaceae family from a number of industrial perspectives. Large-scale cultivation has increased the incidence of various biotic stresses. Diseases such as leaf spot, leaf blight, seedling blight, powdery mildew, and virus infections have been reported in the major growing regions, which has caused significant global yield losses. Genomic and transcriptomic analyses of the castor bean genome have led to the identification of nearly 170 predicted disease resistance genes, including members of the NBS-LRR family, along with over 300 defense-response-associated transcription factors. Furthermore, genetic markers linked to disease resistance genes have also been cataloged and comparative genomics has uncovered common molecular descriptors associated with disease resistance in castor bean and other important members of the Euphorbiaceae family. Resistance genes that are common in Euphorbiaceae can be leveraged for engineering biotic stress tolerance in castor bean. It is anticipated that emerging engineering techniques such as CRISPR/Cas9-based genome editing will enable the enhancement of disease resistance in castor bean. Modern biotechnological approaches and advanced genomic and other "omic" technologies will pave the way to counter disease prevalence and to develop resistance strategies to achieve castor bean's full-yield potential.

6.1 Introduction

In order to generate pathogen-resistant crops that maintain high yields under biotic stress, breeding programs have been focused on the identification of genes that confer resistance to various diseases. Sophisticated and reliable techniques to identify disease resistance genes and to understand their mode of action are required for accelerated breeding. However, high levels of heterozygosity and polyploidy represent serious obstacles for traditional breeding approaches. Modern, genomic-based methodologies are being extensively used for gene identification and molecular cloning. These technologies facilitate mapping of important genes relevant to major traits including disease resistance in numerous

R. S. Chauhan (✉)
Department of Biotechnology, Bennett University, 201310 Greater Noida, Uttar Pradesh, India
e-mail: rajinder.chauhan@bennett.edu.in

Present Address:
A. Sood
Biotechnology Division, CSIR-Institute of Himalayan Bioresource Technology, Post Box No. 6, 176061 Palampur, Himachal Pradesh, India

C. Kole and P. Rabinowicz (eds.), *The Castor Bean Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97280-0_6

plant species (Joyeux et al. 1999; Dracatos et al. 2009; Arafa et al. 2018) without the need of complex procedures such as map-based cloning.

Resistance to specific pathogens and pests in plants is often due to single resistance (*R*) genes (Chauhan and Sood 2013). Many *R* genes have been cloned in numerous plant species, in which they confer resistance to a wide range of pathogens, including bacteria, viruses, and fungi (DeYoung and Innes 2006; McHale et al. 2006; Seo et al. 2006). Although the mechanisms underlying infection vary considerably between pathogens, *R* genes are highly conserved.

There are eight main classes of *R* genes that are summarized in Table 6.1. All these classes contain nucleotide-binding site (NBS), leucine-rich repeat (LRR) domains, and/or associated motifs, except for enzymatic *R* genes.

The majority of *R* genes belong to the NBS-LRR family, which provides resistance against different pathogens such as parasites, bacteria, fungi, viruses, and insects (Dangl and Jones 2001; Martin et al. 2003; Tarr and Alexander 2009; Narusaka et al. 2013). This family is further divided into two subfamilies based on the domains present in their N-terminal regions (Fig. 6.1). One subfamily is called TNL proteins (TIR-NBS-LRR) due to the presence of an N-terminal Toll/interleukin-1 receptor (TIR) motif, while the other is known as CNL (CC-NBS-LRR) because they contain a coiled-coil (CC) N-terminal domain.

NBS-LRR proteins are involved in signal transduction, where the NBS domain binds and hydrolyzes ATP while the LRR domain binds to pathogen-associated proteins. The TIR and CC domains also have major roles in signaling regulation (DeYoung and Innes 2006). NBS domains also have Kinase-1 (P-loop), Kinase-2, Kinase-3, and other short motifs with undefined functions (van der Biezen and Jones 1998). These regions are generally conserved across the plant kingdom, which bring opportunities for the development of cross-species genetic markers to identify resistance genes in other plant species.

The identification and experimental validation of new *R* genes will accelerate genetic improvement programs and breeding practices to develop disease-resistant varieties in important crops. *R* genes, especially NBS-LRR ones, have been actively targeted to develop genetically improved and transgenic-resistant varieties, often by overexpression of the selected *R* gene. These approaches confer resistance specific to different viral, bacterial, fungal, or even insect diseases.

Castor bean (*Ricinus communis*) is a member of the Euphorbiaceae family and is considered an economically important plant owing to its various industrial applications. The industrial relevance of this species has led to its large-scale

Table 6.1 Major categories of resistance (*R*) genes in plants

S. No.	Major *R* gene category	Example	Reference
1	NBS-LRR-TIR	L6, RPP5	Parker et al. (1997)
2	NBS-LRR-CC	RPM1, RPS2	Bisgrove et al. (1994); Mackey et al. (2003)
3	LRR-TrD	Cf-2, Cf-4	Joosten and de Wit (1999)
4	LRR-TrD-Kinase	Xa21	Wang et al. (1998); Gao et al. (2013)
5	TrD-CC	RPW8	Xiao et al. (2001)
6	TIR-NBS-LRR-NLS-WRKY	RRS1R	Deslandes et al. (2002); Zhang et al. (2016)
7	LRR-TrD-PEST-ECS	Ve2	Kawchuk et al. 2001
8	Enzymatic *R* genes	Rpg1	Brueggeman et al. (2002)

CC—coiled-coil; ECS—endocytosis cell signaling domain; LRR—leucine-rich repeats; NBS—nucleotide-binding site; NLS—nuclear localization signal; PEST—amino acid domain; TIR—Toll/interleukin-1 receptor; TrD—transmembrane domain; WRKY—amino acid domain

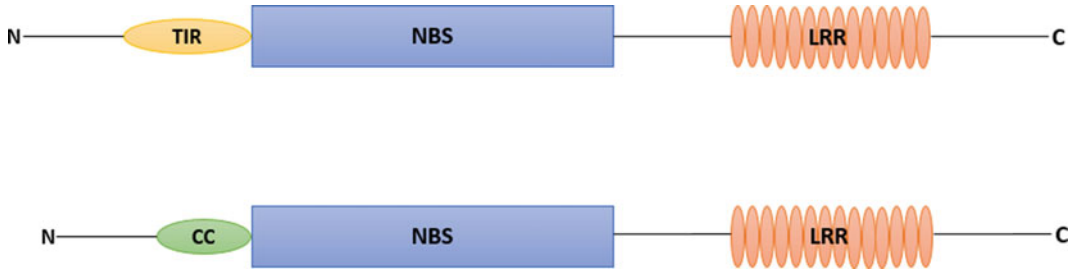

Fig. 6.1 Diagrammatic structure of a NBS-LRR gene (NBS—nucleotide-binding site; LRR—leucine-rich repeat; TIR—Toll and interleukin-1 receptors; CC—coiled-coil; N—amino terminus; C—carboxyl terminus). *Source* McHale et al. (2006)

Table 6.2 Major pathogens responsible for castor bean yield losses

Pathogen	Pathogen type	Disease	Reference
Cercospora coffeicola	Fungi	Leaf spot	Souza and Maffia (2011)
Xanthomonas ricinicola	Bacteria	Leaf blight	Sabet (1959)
Phytophthora palmivora	Fungi	Seed blight	Uchida and Aragaki (1988)
Leveillula taurica	Fungi	Powdery mildew	Mirzaee et al. (2011)
Botryotinia ricini	Fungi	Gray mold	Soares (2012)
Fusarium oxysporum	Fungi	Fusarium wilt	Vahunia et al. (2017)

cultivation, which in turn has increased the incidence of various diseases. Castor bean's vulnerability to various diseases and pathogens can significantly reduce its overall growth and yield. Bacteria and fungi are the most prevalent pathogens affecting castor bean while leaf spot is a major disease in important growing regions such as Costa Rica, India, and Yugoslavia (Table 6.2). The fungus *Cercospora coffeicola* is responsible for leaf spot disease in coffee, and it has been shown to also infect castor bean (Souza and Maffia 2011). Bacterial leaf blight caused by *Xanthomonas ricinicola* results in overall reduction of plant growth and yield losses, and brown/black angular lesions have been prevalent in Sudan and other African countries (Sabet 1959). Another major castor bean pathogen is *Phytophthora palmivora*, which is responsible for seedling blight that causes large yield losses and high mortality rate (Uchida and Aragaki 1988). There have also been reports of powdery mildew disease caused by anamorph stages of the fungus *Leveillula taurica* that devastated a large castor bean population in Iran (Mirzaee et al. 2011).

Viral diseases have also been reported in castor bean. Reddy et al. (2014) reported the presence of tobacco streak virus in the South Indian state of Andhra Pradesh. Symptoms of this viral infection include vein mosaic and necrotic spots on the abaxial surface of leaves. In Greece and Italy, olive latent virus 2 caused leaf yellowish speckling, mottling, and arabesque line patterns resulting in high mortality rate in castor bean (Parrella et al. 2008). In Lucknow, India, Raj et al. (2010) reported significant yield losses in castor bean fields due to cucumber mosaic virus, which showed symptoms like blistering, leaf distortion, and severe mosaic, and recently, Mirhosseini and Nasrollah-Nejad (2017)

described similar symptoms due to infection of castor bean by this virus in some regions of Iran.

There is limited information on the molecular basis of disease resistance in castor bean. The availability of its genome sequence and transcriptome data has opened new paths for detailed analysis and characterization of castor bean resistance genes. The castor bean genome sequence is currently in the form of a draft, composed of nearly 26,000 "scaffolds" or fragments of different sizes, ranging from a few thousand (kbp) to a few million (Mbp) base pairs. Within 69 of those scaffolds, automated gene annotation identified 121 putative disease resistance genes (Chan et al. 2010), which included three classes: NBS-LRR, extracellular leucine-rich repeat (eLRR), and dirigent-like proteins (Fristensky et al. 1988; van Ooijen et al. 2007).

6.2 Disease Resistance Genes in Castor Bean

In addition to the 121 putative disease resistance genes identified in the castor bean genome, Sood et al. (2014) identified 47 novel NBS-LRR resistance genes through transcriptomic analysis. Alignment of these genes to the castor bean genome sequence showed that they had not been previously annotated as such. All the NBS-LRR resistance genes thus identified showed a larger proportion of TNLs than CNLs. The N-terminal domains such as TIR and CC are usually involved in pathogen recognition. In the transcriptome study mentioned above, 318 defense-response-related transcription factors were also identified (Sood et al. 2014). These genes belong to different transcription factor families such as *NAM*, *WRKY, Homeo-domain*, *ERF/AP2-EREBP*, *bZIP*, *Whirly*, and *SBP*. Other genome-wide analyses of castor bean identified multiple copies of AP2/ERF transcription factors, which are involved in biotic and abiotic stress tolerance in dicotyledonous plants (Xu et al. 2013).

Conserved regions in these gene families can be used to identify other potential resistance genes, also called resistance gene analogs (RGAs), by designing cross-species markers using polymerase chain reaction (PCR) with degenerate primers in other plant species. These RGAs are frequently found in the vicinity of major resistance genes or quantitative trait loci (QTL) governing the resistance trait. Thus, RGAs have been used as valuable markers for disease resistance breeding through marker-assisted selection. Gedil et al. (2012) isolated several major RGAs from castor bean and other important members of Euphorbiaceae family. A total of 86 RGAs were identified using sequence homology and degenerate PCR in castor bean, which will help in future characterization and validation of predicted *R* genes, as well as molecular breeding for biotic stress resistance.

Although nearly 170 *R* genes have been identified in castor bean, it is possible that many other *R* genes are yet to be identified, considering that other Euphorbiaceae members have shown larger numbers of *R* genes e.g., approximately 200 in physic nut (Sato et al. 2011), over 300 in cassava (Lozano et al. 2015; Wolfe et al. 2016; Kayondo et al. 2018), and more than 480 in rubber tree (Lau et al. 2016).

The identification of additional *R* genes in castor bean through advanced "omics" approaches will facilitate effective resistance management through biotechnological interventions.

6.3 Comparative Analysis of Castor Bean *R* Genes

Along with castor bean, several other Euphorbiaceae plant species such as physic nut (*Jatropha curcas*), cassava (*Manihot esculenta*), and

rubber tree (*Havea brasiliensis*) are economically important as industrial or staple crops. Castor bean and physic nut are considered promising biofuel crops as well. In spite of these two crops' economic relevance, thorough breeding programs to develop disease-resistant genotypes are still in their infancy. A comparative genomic analysis of castor bean and physic nut showed that these two species share a total of 13,887 gene families (Wu et al. 2015). Predicted homologous NBS-LRR genes found to be conserved in both physic nut and castor bean are potential targets for enhancing biotic stresses resistance not only in these two species but in other Euphorbiaceae members as well (Marone et al. 2013; Song et al. 2013; Wang et al. 2003). This comparative study also identified 70 transcription factors associated with defense response that were common between the two plants, suggesting that defense response mechanisms may be conserved between the two species (Voitsik et al. 2013; Chen et al. 2013).

The genome sequence of castor bean has also been used for cloning and identification of RGAs in *Jatropha integerrima,* another important species in the genus *Jatropha* (Sharma 2011). These RGAs are potential targets to explore the molecular basis of disease resistance within this genus.

Transcriptomic studies aimed at understanding the complex molecular processes involved in disease induction and resistance such as those carried out in physic nut (Sood and Chauhan 2017) will help understand these processes in closely related plants like castor bean.

6.4 Marker Development, Marker-Assisted Selection, and Breeding of Disease Resistance in Castor Bean

Although conventional plant breeding approaches are important and still in use, they are time-consuming and often cannot deliver solutions to pathogen outbreaks. Molecular markers are widely used for crop improvement through genetic mapping and breeding. Marker-assisted selection (MAS) is the application of single marker-trait associations to plant breeding. In MAS, the trait of interest is selected through a linked marker, rather than the gene responsible for that trait. MAS has become the breeding approach of choice for disease resistance in the last few decades. The availability of growing numbers of genome sequences is an important source of data for the development of new markers for fine mapping important traits such as resistance to biotic stresses (Neumann et al. 2011). High-throughput genotyping techniques facilitate tracking introgressed *R* genes in crossing

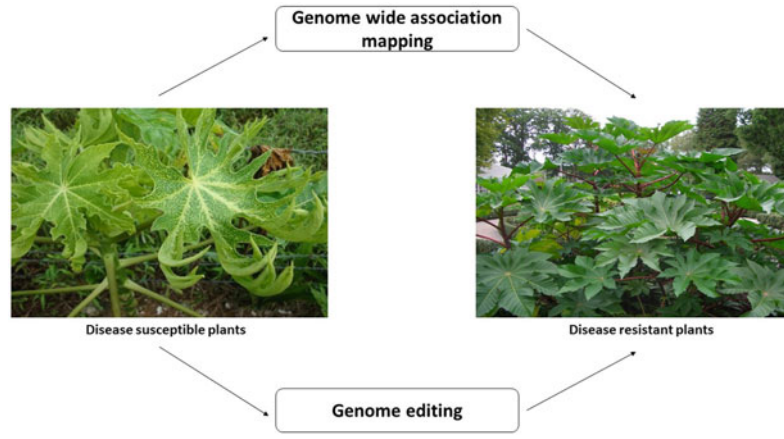

Fig. 6.2 Possible approaches to develop disease-resistant castor bean plants

programs. Many studies using molecular markers to identify *R* genes have been reported, mainly for food crops. However, only a few such studies have been conducted for industry-relevant crops like castor bean. For example, Dhingani et al. (2012) characterized the genetic diversity of fusarium wilt resistance genes in castor bean employing randomly amplified polymorphic DNA (RAPD), inter-simple sequence repeat (ISSR), and simple sequence repeat (SSR) markers, and Reddy et al. (2011) identified two RAPD markers closely linked to the same traits. A combination of MAS and classical breeding approaches was also employed to map such traits, resulting in three RAPD markers linked to the wilt resistant gene in castor bean (Singh et al. 2011). Molecular markers in castor bean have also been used in a genetic diversity analysis that identified natural resistance to fusarium wilt (Anjani 2010). Target region amplification polymorphism (TRAP) markers have also been generated from mRNA sequences to assess genetic diversity of pathogen resistance and other traits among large castor bean populations (Simoes et al. 2017).

6.5 Conclusions

Although the mapping, identification, and characterization of *R* genes in castor bean is only starting, the completion of its genome sequence along with that of other members of the Euphorbiaceae family has paved the way to achieve deeper understanding of the molecular mechanisms underlying biotic stresses in order to generate resistant lines of this important industrial crop. The deployment of state-of-the-art high-throughput approaches as well as modern genetic engineering technologies will allow the introduction of precise genetic alterations into commercial varieties without the need of traditional breeding approaches (Fig. 6.2). Genome engineering technologies like CRISPR/Cas9-based gene editing have been used in different plant species to alter components of their immunity mechanisms, including susceptibility genes (*S*-genes), in order to obtain pathogen resistance. However, single-*R*-gene engineering confers pathogen-specific resistance. The design of synthetic immune receptors with multiple pathogen recognition sites can overcome this limitation and broaden the range of pathogens to which they confer protection. Another strategy to pursue wide pathogen resistance is to introduce various natural or engineered resistance genes within the same cultivar. These precise genome engineering technologies open new avenues to generate desired mutations avoiding unwanted pleiotropic effects.

Genomic-enabled advances in automation will also accelerate marker-assisted selection and pyramiding of resistance traits, as well as quantitative disease resistance loci analysis. The development of new molecular markers and the identification of more disease resistance genes in castor bean will be a key to reduce yield losses due to diseases, allowing the generation of elite lines to maximize production of high-quality specialty oils.

References

Anjani K (2010) Pattern of genetic diversity among Fusarium wilt resistant castor germplasm accessions (*Ricinus communis* L.). Elect J Plant Breed 1(2):182–187

Arafa RA, Soliman NEK, Moussa OM, Kamel SM, Shirasawa K (2018) Characterization of Egyptian *Phytophthora infestans* population using simple sequence repeat markers. J Gen Plant Pathol 84:104–107

Bisgrove SR, Simonich MT, Smith NM, Sattler A, Innes RW (1994) A disease resistance gene in Arabidopsis with specificity for two different pathogen avirulence genes. Plant Cell 6:927–933

Brueggeman R, Rostoks N, Kudrna D, Kolian A, Han F, Chen J, Druka A, Steffenson B, Kleinhofs A (2002) The barley stem rust-resistance gene Rpg1 is a novel disease-resistance gene with homology to receptor kinases. Proc Natl Acad Sci USA 99:9328–9333

Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J et al (2010) Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotechnol 28:951–956

Chauhan RS, Sood A (2013) Comparative genomics in Euphorbiaceae. In: Bahadur B, Sujatha M, Carels N (eds) *Jatropha*: challenges for a new energy crop, vol 2. Springer, New York, pp 351–374

Chen TZ, Lv YD, Zhao TM, Li N, Yang Y, Yu W et al (2013) Comparative transcriptome profiling of a resistant vs susceptible tomato (*Solanum lycopersicum*) cultivar in response to infection by tomato yellow leaf curl virus. PLoS ONE 8:e80816

Dangl JL, Jones JD (2001) Plant pathogens and integrated defence responses to infection. Nature 14;411 (6839):826–833

Deslandes L, Olivier J, Theulieres F, Hirsch J, Feng DX, Bittner-Eddy P, Beynon J, Marco Y (2002) Resistance to *Ralstonia solanacearum* in *Arabidopsis thaliana* is conferred by the recessive RRS1-R gene, a member of a novel family of resistance genes. Proc Natl Acad Sci USA 99:2404–2409

DeYoung BJ, Innes RW (2006) Plant NBS-LRR proteins in pathogen sensing and host defense. Nat Immunol 7 (12):1243–1249

Dhingani RM, Tomar RS, Parakhia MV, Patel SV, Golakiya BA (2012) Analysis of genetic diversity among different *Ricinus communis* genotypes for macrophomina root rot through RAPD and microsatellite markers. Int J Plant Prot 5:1–7

Dracatos PM, Cogan NOI, Sawbridge TI, Gendall AR, Smith KF, Spangenberg GC, Forster JW (2009) Molecular characterisation and genetic mapping of candidate genes for qualitative disease resistance in perennial ryegrass (*Lolium perenne* L.). BMC Plant Biol 9:62

Fristensky B, Horovitz D, Hadwiger L (1988) cDNA sequences for pea disease resistance response genes. Plant Mol Biol 11:713–715

Gao L, Cao Y, Xia Z, Jiang G, Liu G, Zhang W, Zhai W (2013) Do transgenesis and marker-assisted backcross breeding produce substantially equivalent plants?—a comparative study of transgenic and backcross rice carrying bacterial blight resistant gene Xa21. BMC Genom 14:738

Gedil M, Kumar M, Igwe D (2012) Isolation and characterization of resistant gene analogs in cassava, wild *Manihot* species, and castor bean (*Ricinus communis*). Afr J Biotechnol 11:15111–15123

Joosten M, de Wit P (1999) The tomato-*Cladosporium* fulvum interaction: a versatile experimental system to study plant–pathogen interactions. Annu Rev Phytopathol 37:335–367

Joyeux A, Fortin MG, Mayerhofer R, Good AG (1999) Genetic mapping of plant disease resistance gene homologues using a minimal *Brassica napus* L. population. Genome 42:735–743

Kawchuk LM, Hachey J, Lynch DR, Kulcsar F, van Rooijen G, Waterer DR, Robertson A, Kokko E, Byers R, Howard RJ, Fischer R, Prufer D (2001) Tomato ve disease-resistance genes encode cell surface-like receptors. Proc Natl Acad Sci USA 98:6511–6515

Kayondo SI, Carpio DPD, Lozano R, Ozimati A, Wolfe MD, Baguma Y, Gracen V, Samuel O, Ferguson M, Kawuki R, Jannink JL (2018) Genome-wide association mapping and genomic prediction for CBSD resistance in *Manihot esculenta*. Sci Rep 8:1549

Lau NS, Makita Y, Kawashima M, Taylor TD, Kondo S, Othman AS, Shu-Chien AC, Matsui M (2016) The rubber tree genome shows expansion of gene family associated with rubber biosynthesis. Sci Rep 6:28594

Lozano R, Hamblin MT, Prochnik S, Jannink JL (2015) Identification and distribution of the NBS-LRR gene family in the cassava genome. BMC Genom 16: 1–14

Mackey D, Belkhadir Y, Alonso JM, Ecker JR, Dangl JL (2003) Arabidopsis RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance. Cell 112:379–389

Marone D, Russo MA, Laidò G, Leonardis AMD, Mastrangelo AM (2013) Plant nucleotide binding site–leucine-rich repeat (NBS-LRR) genes: active guardians in host defense responses. Int J Mol Sci 14:7302–7326

Martin GB, Bogdanove AJ, Sessa G (2003) Understanding the functions of plant disease resistance proteins. Annu Rev Plant Biol 54:23–61

McHale L, Tan X, Koehl P, Michelmore RW (2006) Plant NBS-LRR proteins: adaptable guards. Genome Biol 7:212

Mirhosseini HA, Nasrollah-Nejad S (2017) First report of cucumber mosaic virus infecting *Ricinus communis* in India. Plant Dis 101(12):2154

Mirzaee MR, Khodaparast MA, Mohseni M, Ramazani SHR, Soltani-Najafabadi M (2011) First record of powdery mildew of castor-oil plant (*Ricinus communis*) caused by the anamorphic stage of *Leveillula taurica* in Iran. Austral Plant Dis Notes 6:36–38

Narusaka M, Kubo Y, Hatakeyama K, Imamura J, Hiroshi E, Nanasato Y, Tabei Y, Takano Y, Shirasu K, Narusaka Y (2013) Interfamily transfer of dual NB-LRR genes confers resistance to multiple pathogens. PLoS ONE 8(2):e55954

Neumann K, Kobiljski B, Denčić S, Varshney RK, Börner A (2011) Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). Mol Breed 27:37–58

Parker JE, Coleman MJ, Szabo V, Frost LN, Schmidt R, van der Biezen EA, Moores T, Dean C, Daniels MJ, Jones JDG (1997) The Arabidopsis downy mildew resistance gene RPP5 shares similarity to the toll and

interleukin-1 receptors with N and L6. Plant Cell 9:879–894

Parrella G, De Stradis A, Vovlas C (2008) First report of olive latent virus 2 in wild castor bean (*Ricinus communis* L.) in Italy. Plant Pathol 57:392

Raj SK, Snehi SK, Gautam KK, Khan MS (2010) First report of association of cucumber mosaic virus with blister and leaf distortion of castor bean (*Ricinus communis* L.) in India. Phytoparasitica 38:283–289

Reddy BBV, Prasanthi L, Sivaprasad Y, Sujitha A, Krishna TG (2014) First report of tobacco streak virus in castor bean. J Plant Pathol 96:431–439

Reddy RN, Sujatha M, Reddy AV, Reddy AP (2011) Inheritance and molecular mapping of wilt resistance gene (s) in castor (*Ricinus communis* L.). Intl J Plant Breed 5:84–87

Sabet KA (1959) Bacterial leaf blight disease of castor. Ann Appl Biol 47:49–55

Sato S, Hirakawa H, Isobe S, Fukai E, Watanabe A, Kato M, Kawashima K, Minami C, Muraki A, Nakazaki N et al (2011) Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. DNA Res 18:65–76

Seo YS, Rojas MR, Lee JY, Lee SW, Jeon JS, Ronald P, Lucas WJ, Gilbertson RL (2006) A viral resistance gene from common bean functions across plant families and is upregulated in a non-virus-specific manner. Proc Natl Acad Sci USA 103:11856–11861

Sharma A (2011) Development of genome resources and their utilization in *Jatropha* through comparative genomics with castor bean. Ph.D. thesis, Jaypee University of Information Technology, Waknaghat

Simões KS, Silva SA, Machado EL, Brasileiro HS (2017) Development of TRAP primers for *Ricinus communis* L. Genet Mol Res 16:1–13

Singh M, Chaudhuri I, Mandal SK, Chaudhuri RK (2011) Development of RAPD markers linked to fusarium wilt resistance gene in castor bean (*Ricinus communis* L.). Genet Eng Biotech J 28:1–9

Soares DJ (2012) Gray mold of castor: a review. In: Cumagun CJR (ed) Plant Pathology. InTech, Rijeka, pp 219–240

Song X, Li Y, Hou X (2013) Genome-wide analysis of the AP2/ERF transcription factor superfamily in Chinese cabbage (*Brassica rapa* ssp. pekinensis). BMC Genom 14:573

Sood A, Chauhan RS (2017) Comparative NGS transcriptomics unravels molecular components associated with mosaic virus infection in a bioenergy plant species, *Jatropha curcas* L. Bioenergy Res 10:129–145

Sood A, Jaiswal V, Chanumolu SK, Malhotra N, Pal T, Chauhan RS (2014) Mining whole genomes and transcriptomes of Jatropha (*Jatropha curcas*) and castor bean (*Ricinus communis*) for NBS-LRR genes and defense response associated transcription factors. Mol Biol Rep 41(11):7683–7695

Souza AGC, Maffia LA (2011) First report of *Cercospora coffeicola* causing cercospora leaf spot of castor beans in Brazil. Plant Dis 95(11):1479

Tarr DE, Alexander HM (2009) TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. BMC Res Notes 2:197

Uchida JY, Aragaki M (1988) Seedling blight of castor bean in Hawaii caused by *Phytophthora palmivora*. Plant Dis 72:1994

Vahunia B, Singh P, Patel NY, Rathava A (2017) Management of fusarium wilt of castor (*Ricinus communis* L.) caused by *Fusarium oxysporum* f. sp. *ricini* with antagonist, botanical extract and pot experiment. Intl J Curr Microbiol App Sci 6(9):390–395

van der Biezen EA, Jones JDG (1998) The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. Curr Biol 8:226–227

van Ooijen G, van den Burg HA, Cornelissen BJ, Takken FL (2007) Structure and function of resistance proteins in solanaceous plants. Annu Rev Phytopathol 45:43–72

Voitsik AM, Muench S, Deising HB, Voll LM (2013) Two recently duplicated maize NAC transcription factor paralogs are induced in response to colletotrichum graminicola infection. BMC Plant Biol 13:85

Wang GL, Ruan DL, Song WY, Sideris S, Chen LL, Pi LY, Zhang SP, Zhang Z, Fauquet C, Gaut BS, Whalen MC, Ronald PC (1998) Xa21D encodes a receptor-like molecule with a leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution. Plant Cell 10:765–779

Wang HZ, Zhao PJ, Xu JC et al (2003) Virus resistance in transgenic watermelon plants containing a WMV-2 coat protein gene. Acta Genet Sin 30:70–75

Wolfe MD, Rabbi IY, Egesi C, Hamblin M, Kawuki R, Kulakow P, Lozano R, Carpio DPD, Ramu P, Jannink JL (2016) Genomewide association and prediction reveals genetic architecture of cassava mosaic disease resistance and prospects for rapid genetic improvement. Plant Genome 9:1–13

Wu P, Zhou C, Cheng S, Wu Z, Lu W, Han J, Chen Y, Chen Y, Ni P, Wang Y, Xu X, Huang Y, Song C, Wang Z, Shi N, Zhang X, Fang X, Yang Q, Jiang H, Chen Y, Li M, Chen F, Wang J, Wu G (2015) Integrated genome sequence and linkage map of physic nut (*Jatropha curcas* L.), a biodiesel plant. Plant J 81(5):810–821

Xiao S, Ellwood S, Calis O, Patrick E, Li T, Coleman M, Turner JG (2001) Broad-spectrum mildew resistance in *Arabidopsis thaliana* mediated by RPW8. Science 291:118–120

Xu W, Li F, Ling L, Liu A (2013) Genomewide survey and expression profiles of the AP2/ERF family in castor bean (*Ricinus communis* L.). BMC Genom 14:785

Zhang C, Chen H, Cai T, Deng Y, Zhuang R, Zhang N, Zeng Y, Zheng Y, Tang R, Pan R, Zhuang W (2016) Overexpression of a novel peanut NBS-LRR gene AhRRS5 enhances disease resistance to *Ralstonia solanacearum* in tobacco. Plant Biotechnol J 15(1): 39–55

Genome Sequence of Castor Bean

7

Agnes P. Chan

Abstract

Castor bean (*Ricinus communis*) was the first plant species with its genome sequenced in the spurge family which also includes cassava, rubber tree, and physic nut. Castor bean is a unique plant species because the seeds are a primary source for the natural product castor oil that contains an unusual fatty acid of economic interest, as well as ricin, the world's deadliest toxin to humans. A draft version of the castor bean genome has been produced using a conventional clone-based and paired-end sequencing approach that reached a sequence coverage of 4.6-fold. Based on the genome sequence assembly, members of the ricin gene family and key oil synthesis genes were identified in a comprehensive manner, representing an important resource for future trait improvement. Importantly, comparative genomics analysis of the castor bean genome with other dicotyledonous plant genomes supported the presence of a conserved ancient hexaploidization event shared throughout the dicot lineage.

7.1 Development and Current Landscape of Plant Genomics

Deciphering the genetic code within the plant domain in the tree of life has greatly improved our understanding of genes, regulatory networks, and metabolic pathways that control plant growth, responses to environmental stress, and the production of economically valuable natural products. The availability of plant whole genome sequences has also provided a comparative genetics framework for the translation of plant genome science into agricultural improvements and efficient production of food crops.

7.1.1 The First Wave

The first plant genome sequence was completed through an international consortium effort for the model flowering plant *Arabidopsis thaliana* almost two decades ago (Arabidopsis Genome Initiative 2000). It was around the same time when the completion of the first draft of the 2.9-gigabase (Gb) human haploid genome was announced (Lander et al. 2001; Venter et al. 2001). The successful demonstrative sequencing and annotation of the 125-megabase (Mb) *A. thaliana* genome was soon followed by genome sequencing projects of agricultural crop plants such as the two widely cultivated rice subspecies *japonica* and *indica* (420 and 466 Mb, respectively) (Goff et al. 2002; Yu et al.

A. P. Chan (✉)
J. Craig Venter Institute, Rockville, MD, USA
e-mail: achan@jcvi.org

C. Kole and P. Rabinowicz (eds.), *The Castor Bean Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97280-0_7

2002), a grape variety derived from the over 2000 years old Pinot Noir lineage (487 Mb) (The French–Italian Public Consortium for Grapevine Genome Characterization 2007), the first commercial virus-resistant SunUp papaya (372 Mb) (Ming et al. 2008), the 2.3-Gb genome of the crop plant maize (Schnable et al. 2009), and the 950-Mb genome of the nitrogen-fixing legume crop soybean (Schmutz et al. 2010). A summary of published plant genome sequencing efforts can be found on the plaBi Database available as timeline (http://www.plabipd.de/timeline_view.ep) and phylogeny views (http://www.plabipd.de/plant_genomes_pa.ep) (Kawakatsu et al. 2016; Bolger et al. 2018).

7.1.2 Sequencing Gigabase-Size Plant Genomes

As of fall 2018, the genomes of over 200 flowering plants have been sequenced since the publication of that of *A. thaliana* (Chen et al. 2018). Technological advances brought about by the second (e.g., massively parallel) and third-generation (e.g., single-molecule real-time) sequencing platforms have resulted in large cost and time reductions in sequencing production, and also importantly increased read lengths compared to the conventional clone-based Sanger sequencing approach. Decoding plant genomes larger than 10 Gb or with high ploidy and repeat content has been made possible as shown by genomes of the ginkgo tree (10.61 Gb) (Guan et al. 2016), the hexaploid bread wheat (15 Gb) (Zimin et al. 2017a; International Wheat Genome Sequencing Consortium (IWGSC) et al. 2018), the Norway spruce (20 Gb) (Nystedt et al. 2013), and the loblolly pine (22 Gb) (Zimin et al. 2017b).

7.1.3 Sequencing the Plant Evolutionary Tree

A number of plant genome sequencing projects carried out by international consortia are now targeting a broad range of species across the plant phylogenetic tree to further explore the plant genome space. The 1000 Plants Initiative (1KP) international consortium (http://www.onekp.com) has sequenced and analyzed transcriptomes from more than 1000 diverse representative species of green plants (Wickett et al. 2014). As a continuation of the 1KP effort, the 10,000 Plants Genomes Project (10KP) has planned to sequence the genomes of 10,000 plant species in over 5 years between 2018 and 2023 (Cheng et al. 2018). The selected genomes will include seed-plants, non-seed plants, green algae, and protists, with a goal to address fundamental questions in basic and applied sciences such as evolution, gene and genome duplications, trait diversification, genotype to phenotype correlations, and potential for production of medicinal compounds and high-value natural products. The 10KP project will be carried out using a "DNA Nanoball" (DNB) platform, a next-generation sequencing technology combined with methodologies for improving long-range genome contiguity (Cheng et al. 2018; Li and Harkess 2018). In addition, the Open Green Genomes Initiative (OGG) (https://jgi.doe.gov/csp-2018-leebens-mack-open-green-genomes-initiative) that will be conducted in collaboration with the DOE Joint Genome Institute (JGI) has planned to select approximately 35 species representing all major evolutionary lineages in the land plant tree of life, to generate high-quality genome assemblies and annotations (Li and Harkess 2018).

7.1.4 From Single to Multiple Reference Genomes

The Human Genome Project was a multi-year effort that delivered the first draft of the human genome sequence in 2001 (Lander et al. 2001; Venter et al. 2001). At the time, the state-of-the-art high-throughput DNA sequencing platform used an automated version of the classical Sanger sequencing approach developed in the 1975 (Sanger and Coulson 1975). New techniques, chemistries, and instruments in sequencing and genomic analysis have evolved at an explosive pace in the past two decades. It

has been estimated that in 2001 the cost of determining the genome sequence of a human individual was $100–300 million, while by 2015 that estimate was reduced to $1500 (https://www.genome.gov/sequencingcosts). Similarly, sequencing efforts for large and complex plant genomes of comparable size to that of humans (e.g., 2.5 Gb for maize) have not only been realized in recent years (Schnable et al. 2009; Jiao et al. 2017b; Springer et al. 2018; Sun et al. 2018), but also expanded to large-scale efforts such as the sequencing of 26 elite accessions for maize breeding improvement (https://nsf.gov/awardsearch/showAward?AWD_ID=1744001), and dozens of sorghum accessions for phenotyping and biofuel purposes (https://arpa-e.energy.gov/?q=programs/terra). The history and advances of next-generation sequencing technologies have been extensively covered by others in several in-depth reviews (Metzker 2010; Morey et al. 2013; Reuter et al. 2015; Goodwin et al. 2016; Heather and Chain 2016; Levy and Myers 2016).

7.2 How Plant Genomes Are Sequenced

7.2.1 General Challenges

Common challenges of plant genome sequencing involve factors such as genome size, content of repetitive sequences, and heterozygosity. *A. thaliana* was selected as the first plant species for genome sequencing because its genome is relatively small (125 Mb) and has a low repetitive DNA content (32%). Further, inbred lines with minimal heterozygosity were readily available (Arabidopsis Genome Initiative 2000; Maumus and Quesneville 2014). In plant genomes, the majority of the repetitive DNA sequences are represented by transposable elements. Unlike vertebrate genomes, in which repetitive DNA content varies from 0.1 to 25% irrespective of genome size, the repetitiveness of plant genomes correlates with increase in genome size (Jiao and Schneeberger 2017). Based on k-mer frequency analysis, which counts all possible substrings of length k, the repeat content of the approximately 0.4 Gb rice, 0.6 Gb sorghum, and 2 Gb maize genomes was estimated to be 20, 45, and 60%, respectively. Sequencing heterozygous individuals (e.g., from outcrossing species) adds the complications of producing different haplotypes during read assembly because of differences between the maternal and paternal copies of the chromosomes. This issue could be overcome if inbred or doubled haploid accessions are available.

Due to the reasons described above, the first steps in sequencing a plant genome de novo are to obtain estimates of its genome size, repeat contents, and level of heterozygosity (Li and Harkess 2018). Genome size information is available from the Kew Plant DNA C-values database, which is maintained by the Royal Botanic Gardens (Gregory et al. 2007). Alternatively, DNA content per cell can be measured by flow cytometry to deduce the size of a plant genome. If preliminary DNA sequence data for the species of interest are available, a bioinformatics approach using k-mer frequency analysis could provide estimations on the expected genome size, genome repetitiveness, and heterozygosity. The genome size can be estimated by the total number of k-mers (the area under the curve in a k-mer frequency plot) divided by sequencing depth, while repetitiveness can be assessed by the length of the k-mer distribution tail, and heterozygosity by the number of peaks (Li and Harkess 2018).

7.2.2 Genome Sequencing Approaches

The workflow of a genome sequencing project would include sequence read production, assembly of reads into long stretches of contiguous sequence or "contigs," and ordering and orienting groups of contigs into scaffolds. For sequence read production, two major approaches could be considered: "short reads" and "long reads." The short-read approach based on massive parallel sequencing has become the standard platform. This approach, which is currently commercialized

by Illumina (and therefore here it will be referred to as "Illumina"), provides high base-calling accuracy and low cost per base. Illumina reads are relatively short (i.e., 150–300 bp) posing a challenge for the assembly of reads into long genomic contigs. As for the long-read approach, there are two main platforms: single-molecule real-time (SMRT) sequencing and nanopore technology. They are marketed by Pacific Biosciences (referred to from now on as "PacBio") and Oxford Nanopore Technologies (from now on called "ONT"), respectively. These technologies produce over kilobase (kb)-long reads, although the raw reads have a higher base-calling error rates, lower throughput, and a higher cost per base than Illumina. Long-read assemblies have been reported for three genomes of three *A. thaliana* relatives (*Euclidium syriacum*, *Conringia planisiliqua*, and the perennial *Arabis alpina*) and individual *A. thaliana* accessions using long-read approaches (PacBio or ONT-based) complemented with a set of highly accurate but short Illumina reads to improve base-calling accuracy (Berlin et al. 2015; Jiao et al. 2017a; Michael et al. 2018). This hybrid approach that combines long and short reads has shown to be successful for sequencing large genomes and it holds promise for future plant genome sequencing efforts.

A brief overview of traditional approaches and next-generation sequencing approaches for generating short reads, long reads, and long-range scaffolding is discussed below.

7.2.2.1 Traditional Approaches

Clone-by-clone approach. Initial efforts to sequencing large eukaryotic genomes such as humans and plants utilized genomic libraries that harbored large DNA inserts. A common type of cloning vector used was the bacterial artificial chromosome (BAC) (Shizuya et al. 1992), which typically contained inserts of 100–200 kb. To provide extensive coverage of the genome, BAC libraries were constructed to contain sufficient number of clones that would cover at least several genome equivalents.

In the clone-by-clone approach, a physical map was needed to identify a set of partially overlapping BAC clones that spanned the chromosome arms. The physical map was built by restriction endonuclease digestion and agarose electrophoresis of all BAC clones to produce restriction fragment profiles also known as fingerprints (Marra et al. 1997). The fingerprints of the BAC clones were then computationally aligned to identify a minimal set of overlapping clones (a "tiling path") aiming to cover the entire genome (Soderlund et al. 1997). The fingerprint-based physical map approach was used by the Arabidopsis Genome Consortium (Marra et al. 1999) and also the Human Genome Project (McPherson et al. 2001).

For the set of BAC clones that were selected for sequencing, each clone was mechanically sheared to construct a shotgun library that would contain randomly cut fragments originated from the BAC clone. Each BAC shotgun library was sequenced to about sixfold read depth and the reads were assembled based on sequence overlaps. Application of the BAC-based "shotgun sequencing" strategy for clones belonging to the tiling path typically would generate a draft genome sequence that covered around 90% of the genome. The remaining 10% of the genome were sequence gaps and could be due to several factors. First, the gap region was not covered by the tiling path of the physical map and therefore no representative BAC clone was selected. Second, the gap region contained repetitive DNA which prevented read assembly. Third, the gap region contained difficult regions for DNA polymerase to read through such as microsatellites that contained arrays of short tandem repeats.

The problematic genomic regions were often validated and finished to improve the quality of the draft genome. Finishing techniques included sequencing of additional BAC clones, amplifying and sequencing problematic regions with specific primers by polymerase chain reaction (PCR), etc. Finishing was the most labor-intensive and expensive phase of the genomic sequencing process and was needed to achieve a high-quality final sequence.

Whole genome shotgun sequencing. An alternative to the clone-by-clone approach was the whole genome shotgun (WGS) sequencing strategy. The castor bean genome was sequenced using the WGS sequencing strategy (Chan et al. 2010). WGS typically involved the construction of individual genomic libraries of different insert sizes. Paired-reads (i.e., both ends of each clone) sequence information and estimated clone sizes were used to assemble the whole genome using computational algorithms. The very first WGS genome project was successfully carried out using the 1.8 Mb small bacterial genome *Haemophilus influenzae* (Fleischmann et al. 1995). The WGS approach was subsequently applied to genomes of increasing sizes such as the 180 Mb genome of the model organism fruit fly *Drosophila melanogaster* (Myers et al. 2000), and eventually the 3000 Mb genomes of human (Venter et al. 2001) and mouse (Venter et al. 2001; Mouse Genome Sequencing Consortium et al. 2002). The WGS approach had also been applied to two rice subspecies (Goff et al. 2002; Yu et al. 2002). One limitation of the WGS approach at the time was the enormous computational capacity required to assemble the WGS sequence reads using de novo assembly algorithms. The resulting WGS-based genome assemblies often contained numerous gaps and the length of the assembled contigs was short. In complex plant genomes, nested repeats are commonly present as long stretches of repetitive DNA sequences in the intergenic regions (San-Miguel et al. 1996; Shirasu et al. 2000; Wicker et al. 2001), and the repetitive elements are highly conserved (e.g., transposons). It was often not feasible to assemble complex plant genomes to the same level of contiguity as in mammalian genomes where repeats were much less conserved and less complex.

Methylation filtration. In animals and plants, extensive DNA methylation silences transposons and repetitive sequences in the heterochromatin. However, a specific differential reduction in DNA cytosine methylation around gene sequences was discovered as a unique property of plant genomes (Rabinowicz et al. 1999). Such differential methylation observed between genes and repetitive DNA has been used to develop a gene enrichment approach called Methylation Filtration (MF). This method allowed the rapid selection and sequencing of genic regions in large and repetitive plant genomes. The technique involves randomly fragmenting plant genomic DNA, cloning the fragments into a plasmid vector, and transforming the library into a *E. coli* host that contains a modified cytosine restriction system (McrBC) (Raleigh 1992). Inside the host, methylated genomic fragments such as highly repetitive transposon sequences would be restricted and selected against, since DNA methylation in plants tends to occur in CG and CNG sequences, which often overlap with the McrBC recognition site (Raleigh 1992). The resulting shotgun plasmid library would be enriched for the hypomethylated genomic fragments which would represent genic regions. An "unfiltered" library would also be produced in parallel using a McrBC-deleted strain to serve as a background control to estimate the extent of gene enrichment achieved in the MF library (Rabinowicz 2003). Selection by MF has been shown to be an efficient strategy for sequencing the gene space of large, repetitive, and complex monocot genomes (Rabinowicz et al. 2005), maize (Palmer et al. 2003; Whitelaw et al. 2003) and sorghum (Bedell et al. 2005). The latest version of MF selection applied to the sugarcane genome demonstrated the feasibility of combining in vitro McrBC digestion with next-generation Illumina short reads sequencing for the de novo assembly of genomic regions within and around genes, including promoters, microRNAs, and introns (Grativol et al. 2014). Since organelle genomes are largely unmethylated, the MF approach has also been applied as a selection method to enrich for organellar sequences such as the chloroplast and mitochondrion genomes as described in Sect. 7.3.6.

7.2.2.2 Illumina

In the Illumina approach, sequencing-by-synthesis reactions take place inside a flowcell (https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html). Briefly, single DNA molecules bind to the solid surface of

the flowcell and are clonally amplified by PCR to generate individual clusters, a process known as bridge amplification. Sequencing of the clusters occurs in a synchronous manner through repeated flow cycles of each of the four nucleotides with reversible dye-terminators. The attainable read length is up to 150 bp in Illumina's high-throughput "HiSeq" platform, and 300 bp for the lower-throughput "MiSeq" platform. The latest Illumina "NovaSeq" platform produces up to 3 terabases (Tb) per run in 45 h, a further increase in sequencing throughput and shortening of production time (https://www.illumina.com/systems/sequencing-platforms/novaseq/specifications.html) relative to previously released devices. The Illumina platform provides a highly competitive cost per base and base-calling accuracy at the expense of read lengths. For plant genome projects, a hybrid strategy is commonly used to combine the benefits of various sequencing platforms. For example, the 844-Mb potato genome was assembled from over 100-fold coverage of Illumina reads from both ends of each DNA fragment (paired-end), 8.5-fold coverage of another short-read technology (Roche 454) and a much smaller fraction of paired-end sequences of large-insert molecular clones (fosmids and BACs) using the traditional Sanger platform (Potato Genome Sequencing Consortium et al. 2011).

7.2.2.3 PacBio

The PacBio platform uses SMRT cells (Eid et al. 2009; Rhoads and Au 2015; Ardui et al. 2018), made up of hundreds of thousands of sequencing units (called zero-mode waveguide, ZMW). Each unit contains a single DNA polymerase molecule immobilized at the bottom, to which a single DNA molecule binds and starts replication by the incorporation of fluorescently labeled nucleotides. During incorporation, distinct light pulse signals are produced by each of the four nucleotides. A "movie" of light pulses is recorded for each ZMW, which are interpreted to obtain base calls. Base modification information such as methylation can also be detected by analyzing variation in the time between base incorporations in the light-pulse movie. The latest PacBio platform "Sequel" produces reads averaging 10–14 kb. The output per Sequel SMRT cell is ~5 Gb. The single-read base-calling error rate is reported to be 13–15%. Error correction by building a consensus of multiple overlapping sequence reads could attain a final accuracy of 99.99% with ~30-fold read depth coverage (https://www.pacb.com/smrt-science/smrt-sequencing/read-lengths) (Ardui et al. 2018).

Because PacBio technology offers the ability to produce long and contiguous genomic sequences and deep sequencing provides improved error rates, whole genome sequencing and assembly has been attempted using exclusively PacBio data. Among the first plant genomes sequenced with this approach were *A. thaliana* (Berlin et al. 2015) and the desiccation-tolerant grass *Oropetium thomaeum* (VanBuren et al. 2015). PacBio de novo assembly of the perennial model plant *Arabis alpina* has provided contiguous spans across repetitive transposable elements extending the genome space (Willing et al. 2015; Jiao and Schneeberger 2017). Three relatives of *A. thaliana* were sequenced and assembled into only a few hundred long contigs using a combination of PacBio long-read data, optical maps (see Sect. 7.2.2.5), and chromosome conformation capture data for genome scaffolding (Jiao et al. 2017a). Although all these genomes are of modest size (between 125 and 375 Mb), PacBio-based assembly has also been applied to larger and more repetitive plant genomes. The maize inbred B73 reference genome was improved using a combination of PacBio data and high-resolution optical mapping. The updated assembly showed a 52-fold increase in contig length and enabled the identification of repetitive transposable elements and their expansions (Jiao et al. 2017b). The sequence of another maize reference genome Mo17 of 2.2 Gb has been reported using a combination of PacBio sequencing and optical mapping (Sun et al. 2018).

7.2.2.4 Oxford Nanopore Technology

Oxford Nanopore Technology (ONT) offers the latest long-read sequencing technology in a palm-sized device (Jain et al. 2016). Commercial

production of the 100-gram sequencer "minION" started in 2015. Because of the compactness and portability of this device, it has been deployed in many field applications, for example, real-time genomic surveillance of the Ebola virus epidemic in West Africa (Hoenen et al. 2016; Quick et al. 2016), in-flight sequencing on the International Space Station (ISS) (Castro-Wallace et al. 2017), and strain-type surveillance of a potentially fatal RNA virus transmitted by mosquitoes in Florida (Russell et al. 2018). In addition, the long and ultra-long-read properties of ONT have demonstrated the capability of de novo sequencing and assembling a human genome from the cell line of the US National Institute for Standards and Technology's (NIST) reference individual NA12878, the assembly and haplotype phasing of the 4 Mb major histocompatibility complex (MHC) locus, and the closing of remaining gaps in the existing human genome sequence. Over half of the bases sequenced are represented in reads longer than 100 kb, and read lengths up to 882 kb have been reported for the human genome assembly (Jain et al. 2018a). ONT data was also used to resolve and assemble long tracts of near-identical tandem repeats in the centromere of the human Y chromosome (Jain et al. 2018b). In terms of sequencing throughput, data generated from a single minION flowcell run in ~48 h can provide genome-scale coverage for the de novo assembly of 100 Mb-size-range genomes such as the model organisms *Caenorhabditis elegans* (Tyson et al. 2018), *D. melanogaster* and additional *Drosophila* species (Miller et al. 2018; Solares et al. 2018), and an *A. thaliana* accession (Michael et al. 2018).

The major distinction of the ONT technology in contrast to other sequencing platforms is that it uses nanoscale channels to read the sequence of bases in a single DNA molecule without the need of enzymatic synthesis or degradation of the DNA. The current generation of the channels is derived from the *E. coli* transmembrane protein CsgG (sigma S-dependent growth subunit G). The CsgG protein in its native form contains a 0.9-nm pore constriction, which is used as a polypeptide secretion channel for the assembly of curli, a bacterial surface protein fiber (Goyal et al. 2014). Inside the minION flowcell, each pore protein is set in an electrically resistant polymer membrane where an electrode records changes in electrical potentials across the membrane. A motor protein (e.g., polymerase or helicase) attached to the pore protein first helps to unwind double-stranded DNA molecules and allows single-stranded DNA molecules to enter the channel at a controlled speed (450 bases of DNA per second per channel). As the DNA molecule traverses through the protein channel, the profiles of the electric current are captured and can be interpreted in real time to generate sequence read data. In the base caller software, raw MinION signals are converted to read sequences using the recurrent neural network (RNN) algorithm specifically developed for base-calling ONT reads (de Lannoy et al. 2017).

ONT can generate reads in the range of kilobases to as long as 2 Mb to date (https://nanoporetech.com/about-us/news/longer-and-longer-dna-sequence-more-two-million-bases-now-achieved-nanopore). The theoretical read length limit is bounded by the intactness and initial size of the input DNA molecules, and the library construction method. In the minION flowcell, an array of thousands of protein channels is built into a synthetic polymer membrane with up to 512 active channels at a given time. The sequencing throughput per flowcell in general is between 5 and 15 Gb over a 48-h run time (depending on the method used to construct the library), with average read lengths around 6–8 kb and the longest reads possibly reaching over 100 kb. Using plant-specific protocols for DNA purification, the per flowcell yield could reach 8.5 Gb with an average read length of 13 kb (Schalamun et al. 2018). The base accuracy of individual single-pass reads is in the range of ~88–95%. It is also worth noting that the use of the minION platform does not require capital investment for the sequencing device but only the consumables, including the flowcells and laboratory reagents needed for quality control of DNA samples and library construction.

ONT has been applied to a number of plant genomes including those of *Solanum pennellii* (tomato) (1–1.1 Gb) (Schmidt et al. 2017), *A. thaliana* (135 Mb) (Michael et al. 2018),

Lactuca sativa (lettuce) (2.6 Gb) (https://nanoporetech.com/resource-centre/tip-iceberg-sequencing-lettuce-genome), a subspecies of *Musa* (banana) (500–600 Mb) (Belser et al. 2018), a subspecies of *Brassica* (e.g., cabbage) (400–500 Mb) (Belser et al. 2018), *Tectona grandis* L. (teak) (465 Mb) (Yasodha et al. 2018), and *Tulipa gesneriana* (tulip) (34 Gb) (https://nanoporetech.com/resource-centre/beauty-and-beast-assembling-tulip-genome). ONT has also been used to generate a single 160 kb contig of the chloroplast genome of *Eucalyptus pauciflora* (snow gum), using high molecular weight DNA prepared from a rapid one-hour DNA extraction protocol (Mayjonade et al. 2016).

7.2.2.5 Getting Long-Range Genome Information

Long-range information across different parts of the genome is important to link genomic contigs and bridge over repetitive regions such as transposable elements which are common in plant genomes. Below is a brief review of long-range technologies to improve the contiguity of the genome assembly.

Optical mapping. The scaffolding of genome contigs can be achieved by whole-genome optical mapping (WGM). An optical map is in essence a high-resolution restriction map of the physical DNA molecules. The method was originally developed in the 1990s for typing yeast strains (Schwartz et al. 1993). Optical maps provide a genomic fingerprint to align, order, and orient in silico digested DNA sequence contigs (reviewed in Chaney et al. 2016). A new generation of single molecule physical maps (e.g., BioNano Genomics) can now be produced from images of linearized, nicked, and fluorescently labeled long DNA strands (Das et al. 2010).

Hi-C. Chromosome conformation capture (3C) allows three-dimensional mapping of chromosomes by identifying their interactions (inside the nucleus), and through the ligation of DNA fragments that are physically close in their natural conformation (Lieberman-Aiden et al. 2009). DNA is cross-linked in vivo to its associated histones, cleaved by restriction enzymes, and isolated from the cells. The ligated DNA fragments are then amplified and sequenced using Illumina, and the read pairs are used for scaffolding. The more recently developed Hi-C approach is an extension of 3C that creates long-range chromosome conformation information (Burton et al. 2013; Selvaraj et al. 2013).

10× Genomics linked reads. In the 10× Genomics approach, the length of input DNA molecules determines the quality of long-range information. Long DNA fragments (up to 100 kb) are first highly diluted into over 10,000 droplet partitions formed using oil/water emulsions and gel beads. After each droplet receives a unique barcode, all droplets are pooled together for Illumina sequencing (Zheng et al. 2016). As homologous DNA fragments are unlikely to be included in the same partition, each individual haplotype will receive its own barcode and can be distinguished. For a highly heterozygous or polyploid individual, 10× Genomics linked reads has the potential to produce a phased genome. Furthermore, a benchmark study comparing laboratory- and computational-based approaches for whole genome phasing of the NIST human reference individual NA12878 showed that the 10× Genomics data produced the lowest switch error rate and outperformed purely computational phasing approaches (Choi et al. 2018).

7.3 The Castor Bean Genome

7.3.1 Economic and Biosafety Importance

Castor bean is of economic importance because of the high content of high-quality oil in its seeds (da Silva Ramos et al. 1984). Castor oil contains 90% ricinoleic acid, an unusual hydroxylated fatty acid that has unique chemical and physical properties, and is used as vital industrial raw material for high-quality lubricants, paints, coatings, plastics, soaps, medications for skin affections, and cosmetics. Castor bean is also of interest from the biosafety point of view. The

seeds of castor bean contain a high amount of ricin, an extremely toxic protein. Ricin is considered one of the deadliest natural poisons when it is delivered intravenously or inhaled as fine particles. A better understanding of the castor bean genome sequence would provide a genetic tool kit for castor oil production without the complications of ricin toxicity, and forensic analysis to genotype and trace plant origins in potential biosecurity scenarios.

The castor bean genome effort began around 2006 and was carried out at the J. Craig Venter Institute (JCVI), formerly The Institute for Genomic Research (TIGR) and the University of Maryland Institute of Genome Science (IGS). The project was funded as part of the National Institute of Health (NIH) Genome Sequencing Center (GSC) project because of the biosecurity concerns of castor bean. Castor bean seeds can be used as a source to extract the unique and highly poisonous toxin ricin. In 2010, the castor bean genome sequence was published as the first genome from the spurge family (Euphorbiaceae) and the 13th angiosperm genome. The genomes of additional Euphorbiaceae species of economic value have been subsequently determined, including the rubber tree that is used to produce rubber and other chemicals (Rahman et al. 2013; Tang et al. 2016), cassava, which is an important source of food starch in developing countries (Prochnik et al. 2012), and physic nut that has potential as for biodiesel production (Sato et al. 2011). Together with castor bean, these crops are valuable resources for evolutionary and comparative genomics efforts for improvement of traits related to the biosynthesis of natural plant products.

7.3.2 Genome Assembly

The draft sequence of the castor bean genome was generated from the *Ricinus communis* Hale cultivar (Chan et al. 2010) using a whole genome shotgun strategy. Genomic DNA was extracted from etiolated seedlings to reduce the proportion of chloroplast DNA. Over 2 million Sanger sequence reads were generated from plasmid and fosmid libraries and assembled using the Celera assembly software (Myers et al. 2000). The genome assembly contained 25,800 scaffolds, spanned 350 Mb, and covered the genome $\sim$4.6 times. The size of genome assembly is consistent with previously reported flow cytometry measurements that estimated the size of the castor bean genome in 320 Mb. Counting contigs longer than 2 kb, the assembly was constituted by 3500 scaffolds, and it spanned 325 Mb. The N50 was 0.56 Mb, which means that 50% of the assembled bases were contained in contigs spanning 0.56 Mb or longer. The repetitive DNA content was estimated to be over 50% and one-third of the repetitive elements were retrotransposons. The most abundant known repetitive sequences are long terminal repeat (LTR) elements of which 22.7% were Gypsy-type and 9.5% Copia-type retrotransposons. Less than 2% of the repetitive elements were DNA transposons.

Annotation of protein-coding genes was performed using predicted models from multiple gene prediction tools [Fgenesh (Salamov and Solovyev 2000), Augustus (Stanke and Waack 2003), GlimmerHMM (Majoros et al. 2004), and SNAP (Korf 2004)], homology searches against rice and *Arabidopsis* protein sequences, a non-redundant amino acid database, and expressed sequence tags (ESTs) collected from the National Center for Biotechnology Information (NCBI) GenBank. To improve gene annotation, over 50,000 ESTs were sequenced from five cDNA libraries generated from mRNA purified from leaves, flowers, roots, and two different seed developmental stages. The cDNA spliced-alignment tool PASA (Program to Assemble Spliced Alignments) (Haas et al. 2003) was used to assemble and consolidate transcript alignments according to exon and intron boundary alignments of EST and other cDNA sequences. The EvidenceModeler tool (Haas et al. 2008) was used to produce consensus gene models from predicted genes, and alignments of

Table 7.1 Summary statistics for the castor bean draft genome assembly and gene annotation

Assembly	
Fold genome coverage	4.59
Number of scaffolds (all)	25,828
Number of scaffolds (>2 kb)	3500
Total span	350.6 Mb
N50 (scaffolds)	496.5 kb
Largest scaffold	4.7 Mb
Average scaffold length	14 kb
Number of contigs	54,000
Largest contig	190 kb
Average contig length	6 kb
N50 (contigs)	21.1 kb
GC content	32.5%
Gene annotation	
Gene models	31,237
Mean gene length	2258.6 bp
Mean coding sequence length	1004.2 bp
Longest gene	15,849 bp
Mean number of exons per gene	4.2
Mean exon length	251 bp
Longest exon	6590 bp

Modified with permission from Chan et al. (2010)

EST, cDNA, and proteins. A total of 31,237 gene models were identified, 58.5% of which could be grouped into 3020 predicted protein families using the TIGR paralogous family pipeline. A summary of the castor bean genome assembly and gene annotation is shown in Table 7.1.

7.3.3 Conserved Genome Triplication

Studying genomic duplications in the castor bean genome provided a valuable opportunity to examine the evolutionary history of the dicotyledonous lineage. At the time of the castor genome project, there was a debate regarding the origin of polyploidy observed in dicots. One model suggested the occurrence of an ancestral hexaploidization event which took place at one point in the evolution of dicots and is still present in most lineages (The French–Italian Public Consortium for Grapevine Genome Characterization 2007). The other model proposed that all dicot genomes share one duplication event and several successive genome duplications occurred separately in each lineage (Velasco et al. 2007).

A comprehensive search for evidence of genomic triplications in the castor bean haploid genome assembly was carried out by performing alignments of all the protein sequences in the genome and identifying reciprocal best matches. The location of those identified paralogous genes was used to determine any conserved blocks of paralogous genes that may be present among the longest genome assembly scaffolds. Using this approach, a total of 17 triplicated regions were identified. One of the 17 regions is shown in Fig. 7.1a. Triplicated regions refer to genomic regions for which two additional paralogous regions exist in the genome. The sequence alignment results suggested that the castor bean genome underwent a hexaploidization event. To compare the triplication of the castor bean genome to ancestral polyploidization events in the dicot lineage, the triplicated regions identified in the castor bean genome were mapped to the genomes of *A. thaliana* (Arabidopsis Genome Initiative 2000), poplar (Tuskan et al. 2006), and grapevine (The French–Italian Public Consortium for Grapevine Genome Characterization 2007). The castor bean paralogous triplet gene blocks displayed one-to-one, one-to-two, one-to-four, and inconclusive relationships, with their corresponding orthologs in grapevine, poplar, *A. thaliana*, and papaya, respectively (Fig. 7.1b–e). These results altogether supported the presence of a hexaploidization event common to all dicots, and that there was one additional genome duplication in poplar, and two further duplications in the *A. thaliana* genome.

7.3.4 The Ricin Toxin Gene Family

The ricin toxin present in castor bean seeds is an important subject for biosecurity research. Ricin is a type 2 ribosome-inactivating protein (RIP), composed of two subunits linked by a disulfide bond: a 32 kDa ricin toxin A chain that harbors

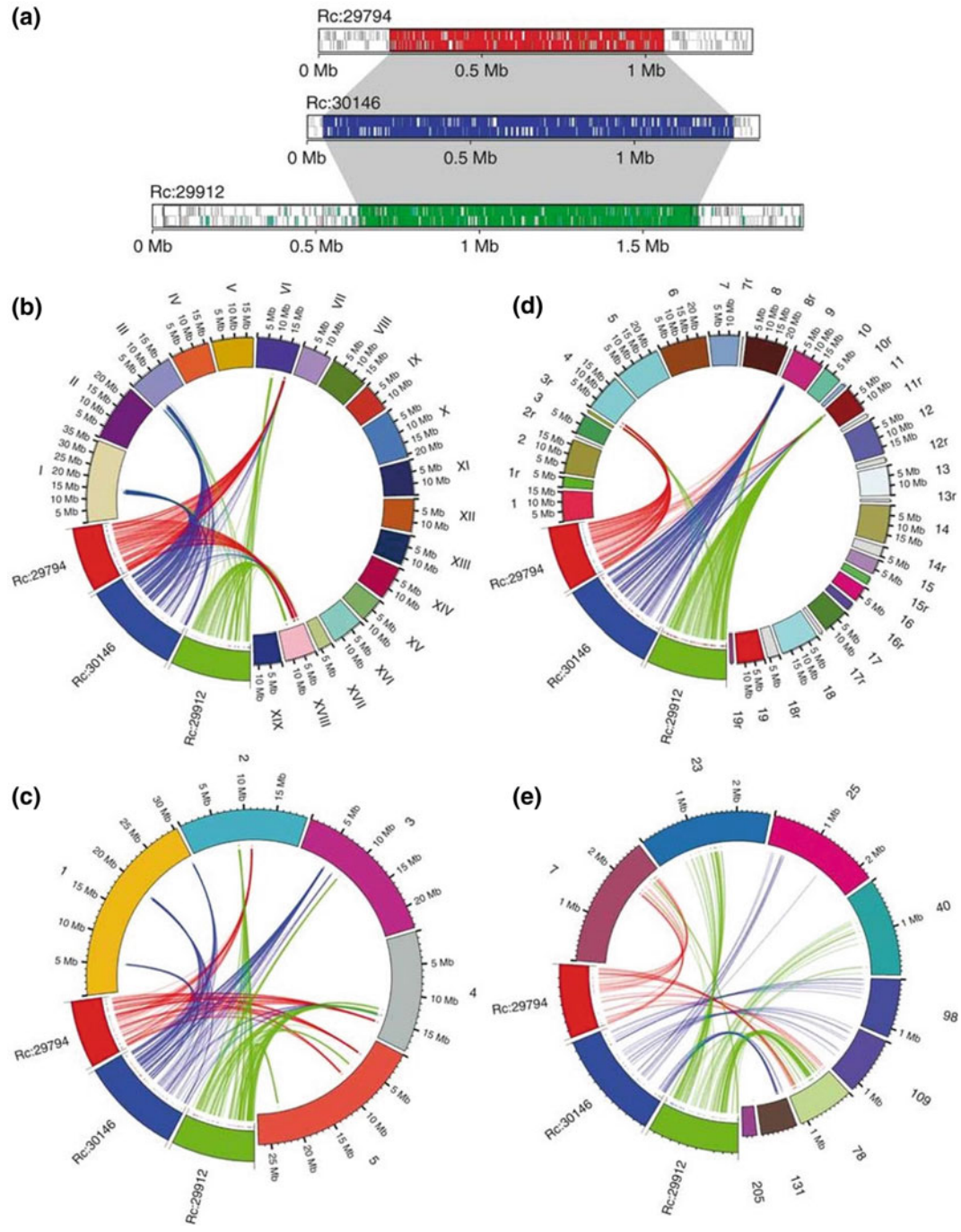

Fig. 7.1 **a** An example of colinearity between three paralogous castor bean genomic regions (triplications). **b**–**e** Putative orthologous gene pairs are shown as colored lines connecting the castor bean scaffolds (noted as Rc: scaffold number) to chromosomes or scaffolds in the other dicot genome. In most cases, one copy of the paralogous castor bean genes corresponds to two genes in poplar (**b**), one gene in grapevine (**c**), and four genes in *A. thaliana* (**d**). The castor bean–papaya relationship (**e**) is inconclusive. Reproduced with permission from Chan et al. (2010)

the ribosome-inactivating activity, and a 34 kDa ricin toxin B chain, with a galactose-binding lectin domain. The ricin toxin A chain is an N-glycosidase that depurinates adenine in a specific residue of the 28S ribosomal RNA and inhibits translation (Endo et al. 1987; Macbeth and Wool 1999). The ricin toxin B chain allows ricin to bind to cell-surface galactosides and enter eukaryotic cells by endocytosis. Early reports based on Southern-blot hybridization suggested the presence of multiple copies (6–8 gene family members) in the castor bean genome (Halling et al. 1985; Tregear and Roberts 1992).

A search for ricin gene sequences in the castor bean genome assembly identified 28 ricin-like gene family members located in 17 genome scaffolds. Some of these ricin-like genes were found tandemly arranged in gene clusters on different genome contigs. The largest gene cluster spans 70 kb and includes a group of five ricin-agglutinin gene family members. Other clusters contain two or three genes in regions ranging between 0.7 and 17 kb. The remaining ten scaffolds contained single gene-family members.

The predicted protein length of the gene family ranged from 66 to 584 amino acids. The shorter probably truncated members could be nonfunctional or pseudogenes. In four cases, the gene sequences were found at the end of a contig or scaffold, which had led to the truncation. In all, 7 of the 28 genes of the gene family encode full-length proteins that contain the RIP and the two lectin domains.

Among the castor bean genome contigs, the precursors for the 577-aa ricin and RCA proteins were identified as gene models RCOM_2159910 (60629.m00002) and RCOM_2160110 (60637.m00004) respectively, in the castor bean genome assembly. As for the seven additional full-length ricin family members, their ribosome inhibiting function had been functionally confirmed using an in vitro transcription/translation system (Leshin et al. 2010).

7.3.5 Oil Metabolism Genes

Castor bean has economic significance as an oilseed crop because of the high oil content in its seeds. Castor bean has evolved two unique features for the biosynthesis of fatty acids (ricinoleic acid) and triacylglycerols (triricinolein) (McKeon et al. 2000). For ricinoleic acid biosynthesis, castor bean possesses an oleic acid hydroxylase (*FAH*) (RCOM_0146820/28035.m000362), which likely evolved from the commonly found Δ12-oleic acid desaturase (*FAD2*) (RCOM_0503360/29613.m000358) (Van De Loo et al. 1995). For the incorporation of ricinoleic acid into oils, there are three diacylglycerol acyltransferases (DGATs) (RCOM_1047540/29912.m005373, RCOM_0613570/29682.m000581 and RCOM_1004000/29889.m003411) (He et al. 2004; Saha et al. 2006; Kroon et al. 2006).

7.3.6 The Castor Bean Organelle Genomes

In addition to the nuclear genome, plants contain plastid and mitochondrion genomes. The chloroplast genomes contain genes essential for structural and functional components of the organelle. Their sequence commonly contains a duplicated region arranged as inverted repeats (IR), a large single-copy (LSC) sequence, and a small single-copy (SSC) sequence (Sugiura 1992; Sugiura et al. 1998). In land plants, chloroplast genome sizes vary between 115 and 165 kb (Palmer 1991). As for the mitochondrion genomes, they are unusually large in higher plants in comparison to the animal counterparts. Plant mitochondrion genomes could span from 208 kb in white mustard (*Brassica hirta*) (Ward et al. 1981; Palmer and Herbon 1987) to 11.3 Mb in the flowering plant *Silene conica* (Sloan et al. 2012), in contrast to 16–20 kb in animals (Boore 1999). Plant mitochondrion genomes typically code for a few RNA genes as well as ribosomal

proteins and subunits of the oxidative phosphorylation complexes and are rich in repeats, including tandem repeats, short repeats, and large repeats. Because of the presence of over 1 kb-long repeats and active homologous recombination, it is often a challenge to determine the exact sequence or structure (circular or linear) of plant mitochondrial DNA (reviewed in Woloszynska 2010).

Classical approaches to obtain the chloroplast and mitochondrion genome sequences have involved cloning fragments from purified organellar DNA obtained either through digestion with restriction enzymes or by random shearing (Sandbrink et al. 1989; Maier et al. 1995; Negruk 2013). In BAC-based genome sequencing projects, organelle genomes are often assembled as by-products by screening BAC clones that contain organelle DNA through PCR or hybridization with organelle-specific probes (Saski et al. 2005).

For the castor bean Hale cultivar, the chloroplast genome sequence was assembled from WGS paired-end Sanger reads derived from the draft genome sequencing project (Rivarola et al. 2011). Sequence reads that showed high similarity to dicot chloroplast sequences were extracted and assembled. Long-range fosmid read pairs were used to resolve the large inverted repeat sequences commonly found in angiosperm chloroplasts. The Hale cultivar chloroplast genome was 163,161 bp with a near-perfect inverted repeat of 27,347 bp.

Using a similar approach, mitochondrion genome sequence reads were extracted from the castor bean genome sequencing project and assembled (Rivarola et al. 2011). The assembled castor bean mitochondrion genome consensus was 502 kb in length, which represents multiple genomic and subgenomic configurations that could be present in vivo. The castor bean mitochondrial genome contained 37 protein-coding and 3 ribosomal RNA genes which are commonly found in other angiosperms.

Taking advantage of the unmethylated nature of chloroplast DNA methylated (McCullough et al. 1992; Fojtová et al. 2001), the MF approach (described above in Sect. 7.2.2.1) was used to sequence and assemble the chloroplast genomes of a number of castor bean accessions retrieved from different locations around the world. MF's selection of hypomethylated sequences can be used to construct total plant DNA libraries enriched in chloroplast sequences. Seven genetically and geographically diverse varieties collected from Ethiopia, India, US Virgin Islands, Puerto Rico, El Salvador, Greece, and Mexico were chosen for chloroplast genome sequencing, single nucleotide polymorphism (SNP) discovery, and phylogenetic studies (Rivarola et al. 2011). An MF library was constructed for each variety and pair-end Sanger reads were generated for several thousand clones per variety. A comparison of the MF library against a regular WGS library showed an approximately 3-fold enrichment in chloroplast sequences. On average, 35% of the MF sequences corresponded to the chloroplast across the seven accessions, with the maximum reaching 58%. In all, MF provided a simple and rapid approach to enrich for not only genic sequences but also plant organelle genomes because it does not require time-consuming plastid preparations, library screenings, or amplification.

The high-quality assembly of the chloroplast genomes sequences allowed the identification of 83 high-quality chloroplast SNPs. Based on the analysis of those SNPs across the chloroplast genomes of the seven cultivars, two main groups of accessions could be identified and grouped into two clades called A and B, which were separated by 69 of 83 SNPs. Several subclades could also be determined within each main clade. Members of clade A included the Hale subclade and the India subclade. The Hale subclade contained the Hale, Puerto Rico, Virgin Islands, and El Salvador accessions. Clade B included the Ethiopia subclade (which included the Mexico variety) and the Greece subclade. The four identified subclades (Hale, India, Greece, and Ethiopia) were used in a large-scale analysis to classify 894 castor bean accessions collected around the world. Genotyping results showed that limited chloroplast genetic diversity exists globally among castor bean germplasm. These results were consistent with the low genetic

variation previously reported for nuclear genomic sequences (Allan et al. 2008; Foster et al. 2010).

7.4 Perspectives

The castor bean genome draft assembly was produced using the classical Sanger sequencing platform (Chan et al. 2010). Further improvements of the completeness and contiguity of the genome assembly would include: (1) closing assembly gaps, linking contigs, identifying possible structural misassemblies, and resolving repetitive regions including the centromeres using long-range data; (2) generating a more complete catalog of coding- and non-coding gene annotation using RNA-seq evidence from spatial and temporal expression datasets; (3) determining the epigenetic landscape of DNA and histone methylation of the genome and their roles in gene regulation. In the short window of the last few decades, technological advances in sequencing production have brought in and established the current mainstream short-read platform (e.g., Illumina) and more recently long-read technologies (e.g., PacBio and ONT). To generate long-range genomic data, the ability of long and ultra-long reads to span kilobase to megabase genomic regions has been demonstrated, which can be used in conjunction with complementary approaches such as whole-genome optical mapping, Hi-C, or 10× Genomics approaches toward achieving a more complete representation of entire chromosome sequences.

The tremendous reduction in sequencing cost has also allowed genomic research to move beyond studying a single reference genome to establishing reference genomes for multiple accessions. In plants, multi-accession sequencing projects have already been carried out in the Arabidopsis 1001 Genomes Project (Weigel and Mott 2009; Kawakatsu et al. 2016), the pearl millet project for 994 lines (Varshney et al. 2017), and the 3000 rice genomes project (3,000 rice genomes project 2014; Wang et al. 2018). These large-scale projects aim at understanding genetic variation, population structure, genomic diversity, and discovering novel alleles and phenotypes to develop a foundational platform for advancing breeding and trait improvement. As for castor bean, there are over 1000 global accessions collected or donated from 48 countries in the USDA germplasm collection (Wang et al. 2011, 2017) which represents a rich resource to apply genomics approaches and tap on novel traits and alleles specific for increased castor oil production, reduced ricin toxin content, and improved adaptabilities to different growing environments and the changing climate for the unique castor bean plant.

References

3,000 rice genomes project (2014) The 3,000 rice genomes project. Gigascience 3:7. https://doi.org/10.1186/2047-217x-3-7

Allan G, Williams A, Rabinowicz PD, Chan AP, Ravel J et al (2008) Worldwide genotyping of castor bean germplasm (*Ricinus communis* L.) using AFLPs and SSRs. Genet Resour Crop Evol 55:365–378. https://doi.org/10.1007/s10722-007-9244-3

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815. https://doi.org/10.1038/35048692

Ardui S, Ameur A, Vermeesch JR, Hestand MS (2018) Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. Nucl Acids Res 46:2159–2168. https://doi.org/10.1093/nar/gky066

Bedell JA, Budiman MA, Nunberg A, Citek RW, Robbins D et al (2005) Sorghum genome sequencing by methylation filtration. PLoS Biol 3:e13. https://doi.org/10.1371/journal.pbio.0030013

Belser C, Istace B, Denis E, Dubarry M, Baurens F-C et al (2018) Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. Nat Plants 4:879–887. https://doi.org/10.1038/s41477-018-0289-4

Berlin K, Koren S, Chin C-S, Drake JP, Landolin JM, Phillippy AM (2015) Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. Nat Biotechnol 33:623–630. https://doi.org/10.1038/nbt.3238

Bolger ME, Arsova B, Usadel B (2018) Plant genome and transcriptome annotations: from misconceptions to simple solutions. Brief Bioinformat 19:437–449. https://doi.org/10.1093/bib/bbw135

Boore JL (1999) Animal mitochondrial genomes. Nucl Acids Res 27:1767–1780

Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO et al (2013) Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat Biotechnol 31:1119–1125. https://doi.org/10.1038/nbt.2727

Castro-Wallace SL, Chiu CY, John KK, Stahl SE, Rubins KH et al (2017) Nanopore DNA sequencing and genome assembly on the international space station. Sci Rep 7:18022. https://doi.org/10.1038/s41598-017-18364-0

Chan AP, Crabtree J, Zhao Q et al (2010) Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotechnol 28:951–956. https://doi.org/10.1038/nbt.1674

Chaney L, Sharp AR, Evans CR, Udall JA (2016) Genome mapping in plant comparative genomics. Trends Plant Sci 21:770–780. https://doi.org/10.1016/j.tplants.2016.05.004

Chen F, Dong W, Zhang J, Lorenzi H, Orvis J et al (2018) The sequenced angiosperm genomes and genome databases. Front Plant Sci 9:418. https://doi.org/10.3389/fpls.2018.00418

Cheng S, Melkonian M, Smith SA, Brockington S, Archibald JM, et al (2018) 10KP: A phylodiverse genome sequencing plan. Gigascience 7. https://doi.org/10.1093/gigascience/giy013

Choi Y, Chan AP, Kirkness E, Telenti A, Schork NJ (2018) Comparison of phasing strategies for whole human genomes. PLoS Genet 14:e1007308. https://doi.org/10.1371/journal.pgen.1007308

da Silva Ramos LC, Tango JS, Savi A, Leal NR (1984) Variability for oil and fatty acid composition in castorbean varieties. J Am Oil Chem Soc 61:1841–1843. https://doi.org/10.1007/BF02540812

Das SK, Austin MD, Akana MC, Deshpande P, Cao H, Xiao M (2010) Single molecule linear analysis of DNA in nano-channel labeled with sequence specific fluorescent probes. Nucl Acids Res 38:e177. https://doi.org/10.1093/nar/gkq673

de Lannoy C, de Ridder D, Risse J (2017) The long reads ahead: de novo genome assembly using the MinION. F1000Res 6:1083. https://doi.org/10.12688/f1000research.12012.2

Eid J, Fehr A, Gray J, Luong K, Lyle J et al (2009) Real-time DNA sequencing from single polymerase molecules. Science 323:133–138. https://doi.org/10.1126/science.1162986

Endo Y, Mitsui K, Motizuki M, Tsurugi K (1987) The mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes. The site and the characteristics of the modification in 28 S ribosomal RNA caused by the toxins. J Biol Chem 262:5908–5912

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF et al (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496–512

Fojtová M, Kovarík A, Matyásek R (2001) Cytosine methylation of plastid genome in higher plants. Fact or artefact? Plant Sci 160:585–593

Foster JT, Allan GJ, Chan AP, Rabinowicz PD, Ravel J et al (2010) Single nucleotide polymorphisms for assessing genetic diversity in castor bean (*Ricinus communis*). BMC Plant Biol 10:13. https://doi.org/10.1186/1471-2229-10-13

Goff SA, Ricke D, Lan T-H, Presting G, Wang R et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92–100. https://doi.org/10.1126/science.1068275

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17:333–351. https://doi.org/10.1038/nrg.2016.49

Goyal P, Krasteva PV, Van Gerven N, Gubellini F, Van den Broeck I et al (2014) Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. Nature 516:250–253. https://doi.org/10.1038/nature13768

Grativol C, Regulski M, Bertalan M, McCombie WR, da Silva FR et al (2014) Sugarcane genome sequencing by methylation filtration provides tools for genomic research in the genus *Saccharum*. Plant J 79:162–172. https://doi.org/10.1111/tpj.12539

Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K et al (2007) Eukaryotic genome size databases. Nucl Acids Res 35:D332–D338. https://doi.org/10.1093/nar/gkl828

Guan R, Zhao Y, Zhang H, Fan G, Liu X et al (2016) Draft genome of the living fossil *Ginkgo biloba*. Gigascience 5:49. https://doi.org/10.1186/s13742-016-0154-1

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr et al (2003) Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucl Acids Res 31:5654–5666

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 spliced alignments. Genome Biol 9:R7. https://doi.org/10.1186/gb-2008-9-1-r7

Halling KC, Halling AC, Murray EE, Ladin BF, Houston LL et al (1985) Genomic cloning and characterization of a ricin gene from *Ricinus communis*. Nucl Acids Res 13:8019–8033

Heather JM, Chain B (2016) The sequence of sequencers: the history of sequencing DNA. Genomics 107:1–8. https://doi.org/10.1016/j.ygeno.2015.11.003

He X, Turner C, Chen GQ, Lin J-T, McKeon TA (2004) Cloning and characterization of a cDNA encoding diacylglycerol acyltransferase from castor bean. Lipids 39:311–318

Hoenen T, Groseth A, Rosenke K, Fischer RJ, Hoenen A et al (2016) Nanopore sequencing as a rapidly deployable ebola outbreak tool. Emerg Infect Dis 22:331–334. https://doi.org/10.3201/eid2202.151796

International Wheat Genome Sequencing Consortium (IWGSC), IWGSC RefSeq principal investigators:, Appels R, Eversole K, Feuillet C et al (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361. https://doi.org/10.1126/science.aar7191

Jain M, Koren S, Miga KH, Quick J, Rand AC et al (2018a) Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol 36:338–345. https://doi.org/10.1038/nbt.4060

Jain M, Olsen HE, Paten B, Akeson M (2016) The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol 17:239. https://doi.org/10.1186/s13059-016-1103-0

Jain M, Olsen HE, Turner DJ, Stoddart D, Bulazel KV et al (2018b) Linear assembly of a human centromere on the Y chromosome. Nat Biotechnol 36:321–323. https://doi.org/10.1038/nbt.4109

Jiao W-B, Accinelli GG, Hartwig B, Kiefer C, Baker D et al (2017a) Improving and correcting the contiguity of long-read genome assemblies of three plant species using optical mapping and chromosome conformation capture data. Genome Res 27:778–786. https://doi.org/10.1101/gr.213652.116

Jiao W-B, Schneeberger K (2017) The impact of third generation genomic technologies on plant genome assembly. Curr Opin Plant Biol 36:64–70. https://doi.org/10.1016/j.pbi.2017.02.002

Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC et al (2017b) Improved maize reference genome with single-molecule technologies. Nature 546:524–527. https://doi.org/10.1038/nature22971

Kawakatsu T, Huang S-SC, Jupe F, Sasaki E, Schmitz RJ et al (2016) Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. Cell 166:492–505. https://doi.org/10.1016/j.cell.2016.06.044

Korf I (2004) Gene finding in novel genomes. BMC Bioinformatics 5:59. https://doi.org/10.1186/1471-2105-5-59

Kroon JTM, Wei W, Simon WJ, Slabas AR (2006) Identification and functional expression of a type 2 acyl-CoA: diacylglycerol acyltransferase (DGAT2) in developing castor bean seeds which has high homology to the major triglyceride biosynthetic enzyme of fungi and animals. Phytochemistry 67:2541–2549. https://doi.org/10.1016/j.phytochem.2006.09.020

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921. https://doi.org/10.1038/35057062

Leshin J, Danielsen M, Credle JJ, Weeks A, O'Connell KP et al (2010) Characterization of ricin toxin family members from *Ricinus communis*. Toxicon 55:658–661. https://doi.org/10.1016/j.toxicon.2009.09.002

Levy SE, Myers RM (2016) Advancements in next-generation sequencing. Annu Rev Genomics Hum Genet 17:95–115. https://doi.org/10.1146/annurev-genom-083115-022413

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. https://doi.org/10.1126/science.1181369

Li F-W, Harkess A (2018) A guide to sequence your favorite plant genomes. Appl Plant Sci 6:e1030. https://doi.org/10.1002/aps3.1030

Macbeth MR, Wool IG (1999) Characterization of in vitro and in vivo mutations in non-conserved nucleotides in the ribosomal RNA recognition domain for the ribotoxins ricin and sarcin and the translation elongation factors. J Mol Biol 285:567–580. https://doi.org/10.1006/jmbi.1998.2337

Maier RM, Neckermann K, Igloi GL, Kössel H (1995) Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. J Mol Biol 251:614–628. https://doi.org/10.1006/jmbi.1995.0460

Majoros WH, Pertea M, Salzberg SL (2004) TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics 20:2878–2879. https://doi.org/10.1093/bioinformatics/bth315

Marra MA, Kucaba TA, Dietrich NL, Green ED, Brownstein B et al (1997) High throughput fingerprint analysis of large-insert clones. Genome Res 7:1072–1084

Marra M, Kucaba T, Sekhon M, Hillier L, Martienssen R et al (1999) A map for sequence analysis of the *Arabidopsis thaliana* genome. Nat Genet 22:265–270. https://doi.org/10.1038/10327

Maumus F, Quesneville H (2014) Deep investigation of *Arabidopsis thaliana* junk DNA reveals a continuum between repetitive elements and genomic dark matter. PLoS ONE 9:e94101. https://doi.org/10.1371/journal.pone.0094101

Mayjonade B, Gouzy J, Donnadieu C, Pouilly N, Marande W et al (2016) Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. Biotechniques 61:203–205. https://doi.org/10.2144/000114460

McCullough AJ, Kangasjarvi J, Gengenbach BG, Jones RJ (1992) Plastid DNA in developing maize endosperm: genome structure, methylation, and transcript accumulation patterns. Plant Physiol 100:958–964

McKeon TA, Chen GQ, Lin JT (2000) Biochemical aspects of castor oil biosynthesis. Biochem Soc Trans 28:972–974

McPherson JD, Marra M, Hillier L, Waterston RH, Chinwalla A et al (2001) A physical map of the human genome. Nature 409:934–941. https://doi.org/10.1038/35057157

Metzker ML (2010) Sequencing technologies—the next generation. Nat Rev Genet 11:31

Michael TP, Jupe F, Bemm F, Motley ST, Sandoval JP et al (2018) High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. Nat Commun 9:541. https://doi.org/10.1038/s41467-018-03016-2

Miller DE, Staber C, Zeitlinger J, Scott Hawley R (2018) High-quality genome assemblies of 15 Drosophila species generated using nanopore sequencing. bioRxiv 267393

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 452:991–996. https://doi.org/10.1038/nature06856

Morey M, Fernández-Marmiesse A, Castiñeiras D, Fraga JM, Couce ML et al (2013) A glimpse into past, present, and future DNA sequencing. Mol Genet Metabol 110:3–24. https://doi.org/10.1016/j.ymgme.2013.04.024

Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, Rogers J et al (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562. https://doi.org/10.1038/nature01262

Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP et al (2000) A whole-genome assembly of Drosophila. Science 287:2196–2204. https://doi.org/10.1126/science.287.5461.2196

Negruk V (2013) Mitochondrial genome sequence of the legume *Vicia faba*. Front Plant Sci 4:128. https://doi.org/10.3389/fpls.2013.00128

Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C et al (2013) The Norway spruce genome sequence and conifer genome evolution. Nature 497:579–584. https://doi.org/10.1038/nature12211

Palmer JD (1991) Plastid chromosomes: structure and evolution. In: Bogorad L, Vasil IK (ed) The molecular biology of plastids. Cell culture and somatic cell genetics of plants, vol 7A. Academic Press, San Diego, pp 5–53

Palmer JD, Herbon LA (1987) Unicircular structure of the *Brassica hirta* mitochondrial genome. Curr Genet 11:565–570

Palmer LE, Rabinowicz PD, O'Shaughnessy AL, Balija VS, Nascimento LU et al (2003) Maize genome sequencing by methylation filtration. Science 302:2115–2117. https://doi.org/10.1126/science.1091265

Potato Genome Sequencing Consortium, Xu X, Pan S, Cheng S, Zhang B et al (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195. https://doi.org/10.1038/nature10158

Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C et al (2012) The Cassava genome: current progress, future directions. Trop Plant Biol 5:88–94. https://doi.org/10.1007/s12042-011-9088-z

Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E et al (2016) Real-time, portable genome sequencing for Ebola surveillance. Nature 530:228–232. https://doi.org/10.1038/nature16996

Rabinowicz PD (2003) Constructing gene-enriched plant genomic libraries using methylation filtration technology. Methods Mol Biol 236:21–36. https://doi.org/10.1385/1-59259-413-1:21

Rabinowicz PD, Citek R, Budiman MA, Nunberg A, Bedell JA et al (2005) Differential methylation of genes and repeats in land plants. Genome Res 15:1431–1440. https://doi.org/10.1101/gr.4100405

Rabinowicz PD, Schutz K, Dedhia N, Yordan C, Parnell LD et al (1999) Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. Nat Genet 23:305–308. https://doi.org/10.1038/15479

Rahman AYA, Usharraj AO, Misra BB, Thottathil GP, Jayasekaran K et al (2013) Draft genome sequence of the rubber tree *Hevea brasiliensis*. BMC Genom 14:75. https://doi.org/10.1186/1471-2164-14-75

Raleigh EA (1992) Organization and function of the mcrBC genes of *Escherichia coli* K-12. Mol Microbiol 6:1079–1086

Reuter JA, Spacek DV, Snyder MP (2015) High-throughput sequencing technologies. Mol Cell 58:586–597. https://doi.org/10.1016/j.molcel.2015.05.004

Rhoads A, Au KF (2015) PacBio sequencing and its applications. Genom Proteom Bioinformat 13:278–289. https://doi.org/10.1016/j.gpb.2015.08.002

Rivarola M, Foster JT, Chan AP, Williams AL, Rice DW et al (2011) Castor bean organelle genome sequencing and worldwide genetic diversity analysis. PLoS ONE 6:e21743. https://doi.org/10.1371/journal.pone.0021743

Russell JA, Campos B, Stone J, Blosser EM, Burkett-Cadena N et al (2018) Unbiased strain-typing of arbovirus directly from mosquitoes using nanopore sequencing: a field-forward biosurveillance protocol. Sci Rep 8:5417. https://doi.org/10.1038/s41598-018-23641-7

Saha S, Enugutti B, Rajakumari S, Rajasekharan R (2006) Cytosolic triacylglycerol biosynthetic pathway in oilseeds. Molecular cloning and expression of peanut cytosolic diacylglycerol acyltransferase. Plant Physiol 141:1533–1543. https://doi.org/10.1104/pp.106.082198

Salamov AA, Solovyev VV (2000) Ab initio gene finding in Drosophila genomic DNA. Genome Res 10:516–522

Sandbrink JM, Vellekoop P, Van Ham R, Van Brederode J (1989) A method for evolutionary studies on RFLP of chloroplast DNA, applicable to a range of plant species. Biochem Syst Ecol 17:45–49. https://doi.org/10.1016/0305-1978(89)90041-0

Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 94:441–448

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D et al (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768

Saski C, Lee S-B, Daniell H, Wood TC, Tomkins J et al (2005) Complete chloroplast genome sequence of Gycine max and comparative analyses with other legume genomes. Plant Mol Biol 59:309–322. https://doi.org/10.1007/s11103-005-8882-0

Sato S, Hirakawa H, Isobe S, Fukai E, Watanabe A et al (2011) Sequence analysis of the genome of an

oil-bearing tree, *Jatropha curcas* L. DNA Res 18:65–76. https://doi.org/10.1093/dnares/dsq030

Schalamun M, Kainer D, Beavan E, Nagar R, Eccles D et al (2018) A comprehensive toolkit to enable MinION long-read sequencing in any laboratory. bioRxiv 289579

Schmidt MH-W, Vogel A, Denton AK, Istace B, Wormit A et al (2017) De novo assembly of a new *Solanum pennellii* accession using nanopore sequencing. Plant Cell 29:2336–2348. https://doi.org/10.1105/tpc.17.00521

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183. https://doi.org/10.1038/nature08670

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115. https://doi.org/10.1126/science.1178534

Schwartz DC, Li X, Hernandez LI, Ramnarain SP, Huff EJ et al (1993) Ordered restriction maps of Saccharomyces *cerevisiae* chromosomes constructed by optical mapping. Science 262:110–114

Selvaraj S, R Dixon J, Bansal V, Ren B (2013) Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing. Nat Biotechnol 31:1111–1118. https://doi.org/10.1038/nbt.2728

Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P (2000) A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res 10:908–915

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T et al (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. Proc Natl Acad Sci USA 89:8794–8797

Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE et al (2012) Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol 10:e1001241. https://doi.org/10.1371/journal.pbio.1001241

Soderlund C, Longden I, Mott R (1997) FPC: a system for building contigs from restriction fingerprinted clones. Comput Appl Biosci 13:523–535

Solares EA, Chakraborty M, Miller DE, Kalsow S, Hall K et al (2018) Rapid low-cost assembly of the *Drosophila melanogaster* reference genome using low-coverage, long-read sequencing. Genes Genome Genet. https://doi.org/10.1534/g3.118.200162

Springer NM, Anderson SN, Andorf CM, Ahern KR, Bai F et al (2018) The maize W22 genome provides a foundation for functional genomics and transposon biology. Nat Genet. https://doi.org/10.1038/s41588-018-0158-0

Stanke M, Waack S (2003) Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19 Suppl 2:ii215–25

Sugiura M (1992) The chloroplast genome. Plant Mol Biol 19:149–168

Sugiura M, Hirose T, Sugita M (1998) Evolution and mechanism of translation in chloroplasts. Annu Rev Genet 32:437–459. https://doi.org/10.1146/annurev.genet.32.1.437

Sun S, Zhou Y, Chen J, Shi J, Zhao H et al (2018) Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. Nat Genet. https://doi.org/10.1038/s41588-018-0182-0

Tang C, Yang M, Fang Y, Luo Y, Gao S et al (2016) The rubber tree genome reveals new insights into rubber production and species adaptation. Nat Plants 2:16073. https://doi.org/10.1038/nplants.2016.73

The French-Italian Public Consortium for Grapevine Genome Characterization (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463. https://doi.org/10.1038/nature06148

Tregear JW, Roberts LM (1992) The lectin gene family of *Ricinus communis*: cloning of a functional ricin gene and three lectin pseudogenes. Plant Mol Biol 18:515–525

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596–1604. https://doi.org/10.1126/science.1128691

Tyson JR, O'Neil NJ, Jain M, Olsen HE, Hieter P et al (2018) MinION-based long-read sequencing and assembly extends the *Caenorhabditis elegans* reference genome. Genome Res 28:266–274. https://doi.org/10.1101/gr.221184.117

VanBuren R, Bryant D, Edger PP, Tang H, Burgess D et al (2015) Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*. Nature 527:508–511. https://doi.org/10.1038/nature15714

Van De Loo FJ, Broun P, Turner S, Somerville C (1995) An oleate 12-hydroxylase from *Ricinus communis* L. is a fatty acyl desaturase homolog. Proc Natl Acad Sci USA 92:6743–6747

Varshney RK, Shi C, Thudi M, Mariac C, Wallace J et al (2017) Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. Nat Biotechnol 35:969–976. https://doi.org/10.1038/nbt.3943

Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A et al (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS ONE 2:e1326. https://doi.org/10.1371/journal.pone.0001326

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ et al (2001) The sequence of the human genome. Science 291:1304–1351. https://doi.org/10.1126/science.1058040

Wang ML, Dzievit M, Chen Z, Morris JB, Norris JE et al (2017) Genetic diversity and population structure of castor (*Ricinus communis* L.) germplasm within the

US collection assessed with EST-SSR markers. Genome 60:193–200. https://doi.org/10.1139/gen-2016-0116

Wang ML, Morris JB, Tonnis B, Pinnow D, Davis J et al (2011) Screening of the entire USDA castor germplasm collection for oil content and fatty acid composition for optimum biodiesel production. J Agri Food Chem 59:9250–9256. https://doi.org/10.1021/jf202949v

Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S et al (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature 557:43–49. https://doi.org/10.1038/s41586-018-0063-9

Ward BL, Anderson RS, Bendich AJ (1981) The mitochondrial genome is large and variable in a family of plants (cucurbitaceae). Cell 25:793–803

Weigel D, Mott R (2009) The 1001 genomes project for *Arabidopsis thaliana*. Genome Biol 10:107. https://doi.org/10.1186/gb-2009-10-5-107

Whitelaw CA, Barbazuk WB, Pertea G, Chan AP, Cheung F et al (2003) Enrichment of gene-coding sequences in maize by genome filtration. Science 302:2118–2120. https://doi.org/10.1126/science.1090047

Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E et al (2001) Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J 26:307–316

Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E et al (2014) Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc Natl Acad Sci USA 111:E4859–E4868. https://doi.org/10.1073/pnas.1323926111

Willing E-M, Rawat V, Mandáková T, Maumus F, James GV et al (2015) Genome expansion of *Arabis alpina* linked with retrotransposition and reduced symmetric DNA methylation. Nat Plants 1:14023. https://doi.org/10.1038/nplants.2014.23

Woloszynska M (2010) Heteroplasmy and stoichiometric complexity of plant mitochondrial genomes—though this be madness, yet there's method in't. J Exp Bot 61:657–671. https://doi.org/10.1093/jxb/erp361

Yasodha R, Vasudeva R, Balakrishnan S, Sakthi AR, Abel N et al (2018) Draft genome of a high value tropical timber tree, Teak (*Tectona grandis* L. f): insights into SSR diversity, phylogeny and conservation. DNA Res. https://doi.org/10.1093/dnares/dsy013

Yu J, Hu S, Wang J, Wong GK-S, Li S et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296:79–92. https://doi.org/10.1126/science.1068037

Zheng GXY, Lau BT, Schnall-Levin M, Jarosz M, Bell JM et al (2016) Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. Nat Biotechnol 34:303–311. https://doi.org/10.1038/nbt.3432

Zimin AV, Puiu D, Hall R, Kingan S, Clavijo BJ et al (2017a) The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. Gigascience 6:1–7. https://doi.org/10.1093/gigascience/gix097

Zimin AV, Stevens KA, Crepeau MW, Puiu D, Wegrzyn JL et al (2017b) An improved assembly of the loblolly pine mega-genome using long-read single-molecule sequencing. Gigascience 6:1–4. https://doi.org/10.1093/gigascience/giw016

Mining Gene Families in the Castor Bean Genome

8

Zhi Zou

Abstract

Castor bean (*Ricinus communis* L.) is a perennial shrub that belongs to the spurge family, Euphorbiaceae. Several characteristics such as low rate of evolution and not having experienced additional whole-genome duplications after the ancient so-called γ event shared by core eudicots make this species an ideal model to investigate lineage-specific evolution of gene families in Euphorbiaceae as well as core eudicots. In this chapter, I will focus on several representative gene families, i.e., AQP, PLCP, Lhc, WRKY, and Dof, which were manually curated.

8.1 Introduction

Castor bean (*Ricinus communis* L., $2n = 20$) is a perennial shrub that belongs to the family Euphorbiaceae, which consists of more than 6000 species also including several economic importance crops such as physic nut (*Jatropha curcas* L.), cassava (*Manihot esculenta* Crantz), and rubber tree (*Hevea brasiliensis* Muell. Arg.) (Zeng et al. 2010; Zou et al. 2016a, 2017b, 2018, 2019a, 2019b, 2019c). Although originated in Africa, castor is now widely cultivated in many tropical, subtropical, and temperate regions (Qiu et al. 2010; Rivarola et al. 2011). Castor produces high level of oil in its seeds, about 90% of which is composed of the unusual hydroxylated fatty acid ricinoleic acid. The nearly uniform ricinoleic acid content of castor oil and the unique chemical properties of this fatty acid make castor bean to be one of the most important non-food oilseed crops for industrial, medicinal, and cosmetic purposes (Ogunniyi 2006).

The draft genome sequence of castor bean was released in 2010, which spans about 400 Mb and is predicted to encode 31,221 protein-coding genes (Chan et al. 2010). Comparative genomic analysis indicated that the diploid castor bean genome did not experience additional whole-genome duplications (WGDs) after the ancient so-called γ hexaploidization event, which was estimated to occur at about 117 million years ago (Mya), shortly before the origin of core eudicots (Chan et al. 2010; Jiao et al. 2012). Along with

Z. Zou (✉)
Key Laboratory of Biology and Genetic Resources of Tropical Crops, Ministry of Agriculture and Rural Affairs, Institute of Tropical Biosciences and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, Hainan, People's Republic of China
e-mail: zouzhi2008@126.com; zouzhi@catas.cn

Z. Zou
Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou 571737, Hainan, People's Republic of China

C. Kole and P. Rabinowicz (eds.), *The Castor Bean Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97280-0_8

its genomic sequence, as of February 2018, several transcriptome analyses have also been carried out in castor bean, resulting in 62,629 Sanger expressed sequence tags (ESTs) and more than 302,253,625 RNA sequencing (RNA-seq) reads (Table 8.1). These datasets provide an opportunity to characterize specific gene families and analyze their evolutionary patterns in Euphorbiaceae and species beyond.

Several genome-wide analyses of castor bean gene families have been reported, e.g., 11S globulin (Chileh et al. 2010), WRKY transcription factor (Li et al. 2012a; Zou et al. 2016b), light-harvesting chlorophyll a/b-binding (Lhc) (Zou et al. 2013), APETALA2/ethylene responsive factor (AP2/ERF) transcription factor (Xu et al. 2013), basic leucine zipper (bZIP) transcription factor (Jin et al. 2014a), DNA binding with one finger (Dof) transcription factor (Jin et al. 2014b; Zou et al. 2019c), nucleotide binding site-leucine-rich repeats (NBS-LRR) (Sood et al. 2014), SQUAMOSA promoter binding protein (SBP)-box transcription

Table 8.1 Overview of castor bean transcriptome data available in NCBI

Tissue/developing stage	Accession no.	EST/read	Instrument	References
Developing endosperm	–	158	Sanger	Van De Loo et al. (1995)
Phloem	–	4720	Sanger	Doering-Saad et al. (2006)
Developing endosperm	–	329	Sanger	Lu et al. (2007)
Developing seed	–	4902	Sanger	Published only in NCBI
Leaf	–	10,346	Sanger	Published only in NCBI
Flower	–	5619	Sanger	Published only in NCBI
Developing seed (12–33 days after pollination)	–	7645	Sanger	Published only in NCBI
Developing seed (40–61 days after pollination)	–	3988	Sanger	Published only in NCBI
Root	–	24,567	Sanger	Published only in NCBI
Others	–	355	Sanger	Published only in NCBI
Developing endosperm II/III	ERX021375	21,858,754	Illumina GA IIx	Brown et al. (2012)
Developing endosperm V/VI	ERX021376	14,326,641	Illumina GA IIx	Brown et al. (2012)
Germinating seed	ERX021377	17,374,537	Illumina GA IIx	Brown et al. (2012)
Leaf	ERX021378	19,322,724	Illumina GA IIx	Brown et al. (2012)
Male developing flower	ERX021379	17,869,625	Illumina GA IIx	Brown et al. (2012)
Seed and endosperm	SRX485027	76,310,445	Illumina HiSeq 2000	Xu et al. (2014)
Stage III endosperm	SRX007402	137,839	454 GS FLX	Troncoso-Ponce et al. (2011)
Stage IV endosperm	SRX007403	159,209	454 GS FLX	Troncoso-Ponce et al. (2011)
Stage IV endosperm	SRX007403	144,142	454 GS FLX	Troncoso-Ponce et al. (2011)
Stage IV endosperm	SRX007404	89,386	454 GS FLX	Troncoso-Ponce et al. (2011)
Stage V endosperm	SRX007405	82,205	454 GS FLX	Troncoso-Ponce et al. (2011)
Stage VI endosperm	SRX007406	170,648	454 GS FLX	Troncoso-Ponce et al. (2011)
Stage VII–VIII endosperm	SRX007407	160,205	454 GS FLX	Troncoso-Ponce et al. (2011)
Stage III–V endosperm	SRX007408	195,250	454 GS FLX	Troncoso-Ponce et al. (2011)
Root	SRX1054812	2,651,111	Ion Torrent PGM	Published only in NCBI
Leaf, root, seed, and endosperm	SRX343933	24,004,526	Illumina HiSeq 2000	Xu et al. (2013)
RML	SRX1330838	17,819,752	Illumina HiSeq 2000	Tan et al. (2016)
ABPL2	SRX1332476	18,001,231	Illumina HiSeq 2000	Tan et al. (2016)
ABML2	SRX1332477	18,014,510	Illumina HiSeq 2000	Tan et al. (2016)
BML1	SRX1332478	17,759,387	Illumina HiSeq 2000	Tan et al. (2016)
ABPL1	SRX1332479	17,665,132	Illumina HiSeq 2000	Tan et al. (2016)
RPL	SRX1332480	18,136,366	Illumina HiSeq 2000	Tan et al. (2016)
Total		302,316,254		

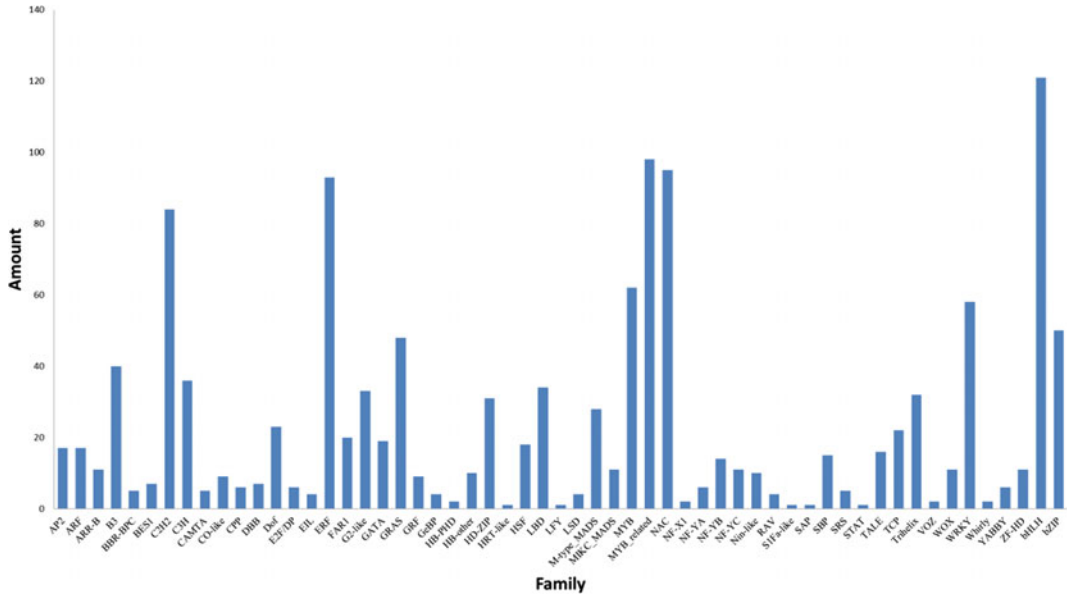

Fig. 8.1 Transcription factor family genes identified in castor bean (data from PlantTFDB 4.0, where the AP2/ERF family were defined as two different families, i.e. AP2 and ERF)

factor (Zhang and Ling 2014), aquaporin (AQP) (Zou et al. 2015b), GRAS transcription factor (GAI, RGA, and SCR) (Xu et al. 2016), and papain-like cysteine protease (PLCP) (Zou et al. 2018). Moreover, 1299 transcription factor genes representing 57 families were also reported in PlantTFDB 4.0 (Fig. 8.1), which were based on the annotated castor bean proteome (Jin et al. 2017). In this chapter, I will focus on several representative gene families, i.e., AQP, PLCP, Lhc, WRKY, and Dof, which were manually curated.

8.2 AQP Gene Family

Aquaporins (AQPs) are a special class of integral membrane proteins that belong to the ancient major intrinsic protein (MIP) superfamily (Maurel et al. 2008). Although they initially raised considerable interest due to their high permeability to water, an increasing body of evidence has shown that some of them also transport certain small molecules such as glycerol, urea, boric acid, silicic acid, arsenic, ammonia, carbon dioxide, oxygen, and hydrogen peroxide (Maurel et al. 2008; Gomes et al. 2009; Zwiazek et al. 2017). Since their first discovery in 1990s (Preston et al. 1992), AQPs have been found in all types of organisms (Gomes et al. 2009; Abascal et al. 2014). Although the overall sequence similarity between different family members can be low, AQPs are characterized by the presence of six transmembrane helices (TM1–TM6) connected by five loops (LA–LE), as well as two highly conserved asparagine–proline–alanine (NPA) motifs located at the N-termini of two half helices (HB and HE) in loops LB and LE. NPA motifs create an electrostatic repulsion of protons and act as a size barrier from one selectivity region of the AQP pore. Another region called the aromatic/arginine (ar/R) selectivity filter causes size and hydrophobicity diversity in the pore constriction site, determining substrate specificity (Törnroth-Horsefield et al. 2006). Genome-wide surveys showed that AQPs are relatively more abundant and diverse in terrestrial plants (Table 8.2) than in microbes and animals. Based on sequence similarity, plant AQPs can be divided into seven main subfamilies: plasma membrane intrinsic proteins (PIPs), tonoplast intrinsic proteins (TIPs), NOD26-like intrinsic proteins (NIPs), small basic intrinsic proteins (SIPs), uncategorized X intrinsic proteins (XIPs), GlpF-like intrinsic proteins (GIPs), and hybrid intrinsic

Table 8.2 Diversity of AQP family genes in high plants (after Zou et al. 2016a)

Species	Common name	Family	Type of organism	PIPs	TIPs	NIPs	SIPs	XIPs	Total
Oryza sativa	Rice	Poaceae	Monocot	11	10	10	2	0	33
Zea mays	Maize	Poaceae	Monocot	13	12	5	3	0	33
Hordeum vulgare	Barley	Poaceae	Monocot	20	11	8	1	0	40
Panicum virgatum	Switchgrass	Poaceae	Monocot	21	20	23	4	0	68
Setaria italica	Foxtail millet	Poaceae	Monocot	12	15	12	3	0	42
Sorghum bicolor	Sorghum	Poaceae	Monocot	13	11	11	3	0	38
Brachypodium distachyon	Common wheat	Poaceae	Monocot	10	10	7	1	0	28
Musa acuminate	Banana	Musaceae	Monocot	18	17	9	3	0	47
Arabidopsis thaliana	Arabidopsis	Brassicaceae	Dicot	13	10	9	3	0	35
Brassica rapa	Chinese cabbage	Brassicaceae	Dicot	23	16	15	6	0	60
Brassica oleracea	Cabbage	Brassicaceae	Dicot	25	19	17	6	0	67
Solanum tuberosum	Potato	Solanaceae	Dicot	15	11	10	3	8	47
Solanum lycopersicum	Garden tomato	Solanaceae	Dicot	14	11	12	4	6	47
Cicer arietinum	Chickpea	Fabaceae	Dicot	8	13	15	4	0	40
Glycine max	Soybean	Fabaceae	Dicot	22	23	17	8	2	72
Gossypium hirsutum	Upland cotton	Malvaceae	Dicot	28	23	12	7	1	71
Vitis vinifera	Grapevine	Vitaceae	Dicot	8	11	9	2	2	32
Citrus sinensis	Sweet orange	Rutaceae	Dicot	8	11	9	3	3	34
Phaseolus vulgaris	Common bean	Fabaceae	Dicot	12	13	10	4	2	41
Linum usitatissimum	Flax	Linaceae	Dicot	16	17	13	2	3	51
Jatropha curcas	Physic nut	Euphorbiaceae	Dicot	9	9	8	4	2	32
Ricinus communis	Castor bean	Euphorbiaceae	Dicot	10	9	8	4	6	37
Manihot esculenta	Cassava	Euphorbiaceae	Dicot	14	13	9	4	2	42
Hevea brasiliensis	Rubber tree	Euphorbiaceae	Dicot	15	17	9	4	6	51
Populus trichocarpa	Poplar	Salicaceae	Dicot	15	17	11	6	6	55

proteins (HIPs) (Anderberg et al. 2011, 2012; Zou et al. 2015a, b, 2016a). The former four subfamilies are widely distributed among plant lineages, whereas XIPs are found in mosses as well as dicots (excluding the Brassicaceae family), and GIPs and HIPs have only been reported in algae and mosses (Danielson and Johanson 2008; Anderberg et al. 2011, 2012; Zou et al. 2016a).

The search of the castor bean genome resulted in 37 loci encoding AQPs (*RcAQPs*: *Ricinus communis AQPs*) (Zou et al. 2015b), corresponding 36 loci reported in the original genome annotation (Chan et al. 2010). Among them, the locus 28747.t000001 was reported to encode a 243 amino acid protein (28747.m000131). However, mRNA sequence alignments revealed that this 11,054-bp sequence is actually containing two genes, and they were denoted as *RcXIP2;1* and *RcXIP3;1* (Zou et al. 2015b). Both of them harbor two introns and putatively encode

306 and 304 residues, respectively (Table 8.3). Moreover, based on a manual revision of their gene structures via aligning of mRNA sequences (including Sanger ESTs and RNA-seq reads) to AQP-encoding scaffolds, the gene models of four other loci (i.e., 28962.t000006, 30101.t000004, 29816.t000013, and 28846.t000001) were also optimized. Homology searches showed that 25 out of the 37 RcAQP genes had matched ESTs in GenBank. Moreover, the expression of the remaining 12 RcAQP genes were supported by RNA-seq reads. In addition, alternative splicing isoforms existing in nine RcAQP-encoding loci were supported by the presence of different mRNA sequences that could be aligned to the same loci (Zou et al. 2015b). Although the current castor bean genome sequence is comprised of 25,763 scaffolds that have not been anchored to the physical chromosomes (Chan et al. 2010), eight scaffolds were shown to harbor two AQP-encoding loci and the other 21 scaffolds contain only one (Table 8.3).

Phylogenetic analysis divides 37 RcAQPs into five subfamilies: PIP, TIP, NIP, XIP, and SIP, with ten, nine, eight, six, or four members, respectively. The PIP subfamily can be further divided into two phylogenetic subgroups (5 RcPIP1s and 5 RcPIP2s), the TIP subfamily into five subgroups (4 RcTIP1s, 2 RcTIP2s, 1 RcTIP3, 1 RcTIP4, and 1 RcTIP5), the NIP subfamily into seven subgroups (1 RcNIP1, 1 RcNIP2, 1 RcNIP3, 2 RcNIP4s, 1 RcNIP5, 1 RcNIP6, and 1 RcNIP7), the SIP subfamily into two subgroups (3 RcSIP1s and 1 RcSIP2), and the XIP subfamily into three subgroups (4 RcXIP1s, 1 RcXIP2, and 1 RcXIP3). A few gene pairs (paralogs) were identified in castor bean, contrasting with a larger number of AQP paralogous gene pairs found in poplar and *Arabidopsis* (Fig. 8.2), which is probably due to recent WGDs reported in those two species (Bowers et al. 2003; Tuskan et al. 2006). Among four castor bean gene pairs identified, they were found as simple (head-to-tail) tandem duplications within the same scaffold (*RcPIP1;2*/*RcPIP1;3* on scaffold29669, *RcSIP1;2*/*RcSIP1;3* on scaffold30010, *RcXIP1;1*/*RcXIP1;4* on scaffold28929, and *RcXIP1;2*/*RcXIP1;3* on scaffold28846), and the intergenic spacer was about 3.1, 0.5, 8.5, or 31.1 kb, respectively (Zou et al. 2015b). Nevertheless, the evolutionary relationships between *RcXIP1;1*/*RcXIP1;4* and *RcXIP1;2*/*RcXIP1;3* were not well resolved. The coding sequence (CDS) of *RcXIP1;2* showed a sequence identity of 85.1% with *RcXIP1;1* and 72.1% with *RcXIP1;3*, while *RcXIP1;1* only exhibited 47.3% identity with *RcXIP1;4*. When using a substitution rate (per base per year) of 6.5×10^{-9} (Gaut et al. 1996), the estimated divergence time between *RcXIP1;1* and *RcXIP1;2* was 30.89 Mya, which is relatively younger than *RcXIP1;2*/*RcXIP1;3* (59.61 Mya) and *RcXIP1;1*/*RcXIP1;4* (74.57 Mya). These results suggest that *RcXIP1;1* and *RcXIP1;2* are proximal or dispersed duplicates, which was defined as separated duplicates within or between chromosomes, respectively. The Ka/Ks ratios of these duplications are all smaller than 1 (from 0.01 to 0.32), indicating that the divergence was driven by purifying selection.

Gene structure analysis revealed that exon–intron structure is usually conserved within a subfamily but diverges between different subfamilies: All PIP subfamily members feature three introns except for *RcPIP2;5*, which contains an additional small intron close to its 5′ end; the TIP subfamily has two introns except for *RcTIP1;1* and *RcTIP1;4* that harbor a single one; the NIP subfamily shows four introns except for *RcNIP5;1* that contains only three; the SIP subfamily features two introns except for *RcSIP1;2* and *RcSIP1;3* that lack introns; and the XIP subfamily features one (XIP1 subgroup) or two (XIP2 and XIP3 subgroups) introns (Table 8.3).

Sequence analysis showed that the 37 deduced RcAQP protein sequences consist of 208–323 amino acids, with a theoretical molecular weight of 22.49–34.93 kDa and a *p*I value of 4.93–9.97. Sequence comparison revealed high levels of diversity within and between the five subfamilies. Topological analysis using TOPCONS (http://topcons.net/) showed that almost all RcAQPs were predicted to harbor six TM helices, except for RcXIP1;4, which was shown to contain only four. The presence of six TM

Table 8.3 List of 37 AQP genes identified in castor (after Zou et al. 2015b)

Name	Locus ID	Transcript ID	Scaffold location	Nucleotide length (bp, from start to stop codons)		Intron no.	EST no.	Alternative splicing	Comment
				CDS	Gene				
RcPIP1;1	29969.t000006	29969.m000266	29969:73,604-75299	867	1370	3	0	–	–
RcPIP1;2	29669.t000017	29669.m000808	29669:104366-101932	864	1304	3	18	–	–
RcPIP1;3	29669.t000018	29669.m000809	29669:109461-107553	867	1243	3	14	–	–
RcPIP1;4	30190.t000465	30190.m011229	30190:2676127-2679732	864	2648	3	109	Yes	–
RcPIP1;5	30174.t000011	30174.m008614	30174:1747452-1750093	861	1712	3	10	–	–
RcPIP2;1	30078.t000130	30078.m002337	30078:846884-849450	867	1185	3	155	Yes	–
RcPIP2;2	27516.t000007	27516.m000174	27516:79695-83007	852	2665	3	14	Yes	–
RcPIP2;3	28076.t000002	28076.m000411	28076:21259-23855	864	1939	3	2	–	–
RcPIP2;4	29869.t000059	29869.m001194	29869:701298-698970	843	1221	3	35	Yes	–
RcPIP2;5	28962.t000006	28962.m000437	28962:25344-23278	813	1654	4	3	Yes	Misannotated
RcTIP1;1	30078.t000047	30078.m002254	30078:302469-304194	756	871	1	307	–	–
RcTIP1;2	28180.t000017	28180.m000392	28180:101668-103298	759	1149	2	5	–	–
RcTIP1;3	29788.t000021	29788.m000338	29788:135388-137224	759	1070	2	60	–	–
RcTIP1;4	29589.t000040	29589.m001261	29589:301877-300492	759	1129	1	7	–	–
RcTIP2;1	30146.t000119	30146.m003542	30146:893065-891669	747	1300	2	47	Yes	–
RcTIP2;2	30101.t000004	30101.m000372	30101:54578-56981	753	1059	2	196	Yes	Misassembled
RcTIP3;1	29681.t000071	29681.m001366	29681:356430-358852	768	994	2	68	Yes	
RcTIP4;1	29794.t000118	29794.m003419	29794:784479-783045	744	1017	2	1	–	–
RcTIP5;1	30147.t000502	30147.m014231	30147:1083709-1086062	759	1241	2	0	–	–
RcNIP1;1	30026.t000052	30026.m001488	30026:382992-385763	816	2079	4	1	–	–
RcNIP2;1	27860.t000001	27860.m000040	27860:3885-7450	894	2598	4	0	–	–
RcNIP3;1	29908.t000084	29908.m006033	29908:1225139-1222967	849	1310	4	0	–	–

(continued)

Table 8.3 (continued)

Name	Locus ID	Transcript ID	Scaffold location	Nucleotide length (bp, from start to stop codons)		Intron no.	EST no.	Alternative splicing	Comment
				CDS	Gene				
RcNIP4;1	29816.t000013	29816.m000676	29816:132884-130426	810	1425	4	0	–	Misannotated
RcNIP4;2	28827.t000002	28827.m000171	28827:15568-18585	759	3018	4	0	–	–
RcNIP5;1	30068.t000130	30068.m002640	30068:792230-798545	897	4934	3	62	–	–
RcNIP6;1	29588.t000015	29588.m000860	29588:109896-113370	927	3320	4	1	–	–
RcNIP7;1	29844.t000176	29844.m003330	29844:993874-996212	897	1258	4	0	–	–
RcXIP1;1	28929.t000003	28929.m000055	28929:27059-25434	930	1195	1	20	–	–
RcXIP1;2	28846.t000001	28846.m000048	28846:5267-3386	912	1346	1	12	–	–
RcXIP1;3	28846.t000002	28846.m000049	28846:38157-36376	972	1913	1	3	–	–
RcXIP1;4	28929.t000002	28929.m000054	28929:16870-16144	627	727	1	0	–	–
RcXIP2;1	28747.t000001	28747.m000131	28747:54568-52108	921	1302	2	18	Yes	Misannotated
RcXIP3;1	28747.t000001	28747.m000131	28747:44459-42629	915	1827	2	0	–	Misannotated
RcSIP1;1	29950.t000061	29950.m001177	29950:345535-339806	720	4822	2	4	–	–
RcSIP1;2	30010.t000008	30010.m000658	30010:76324-74304	705	705	0	0	–	–
RcSIP1;3	30010.t000009	30010.m000659	30010:78002-76765	720	720	0	0	–	–
RcSIP2;1	30045.t000019	30045.m000487	30045:458594-454223	723	3962	2	0	–	–

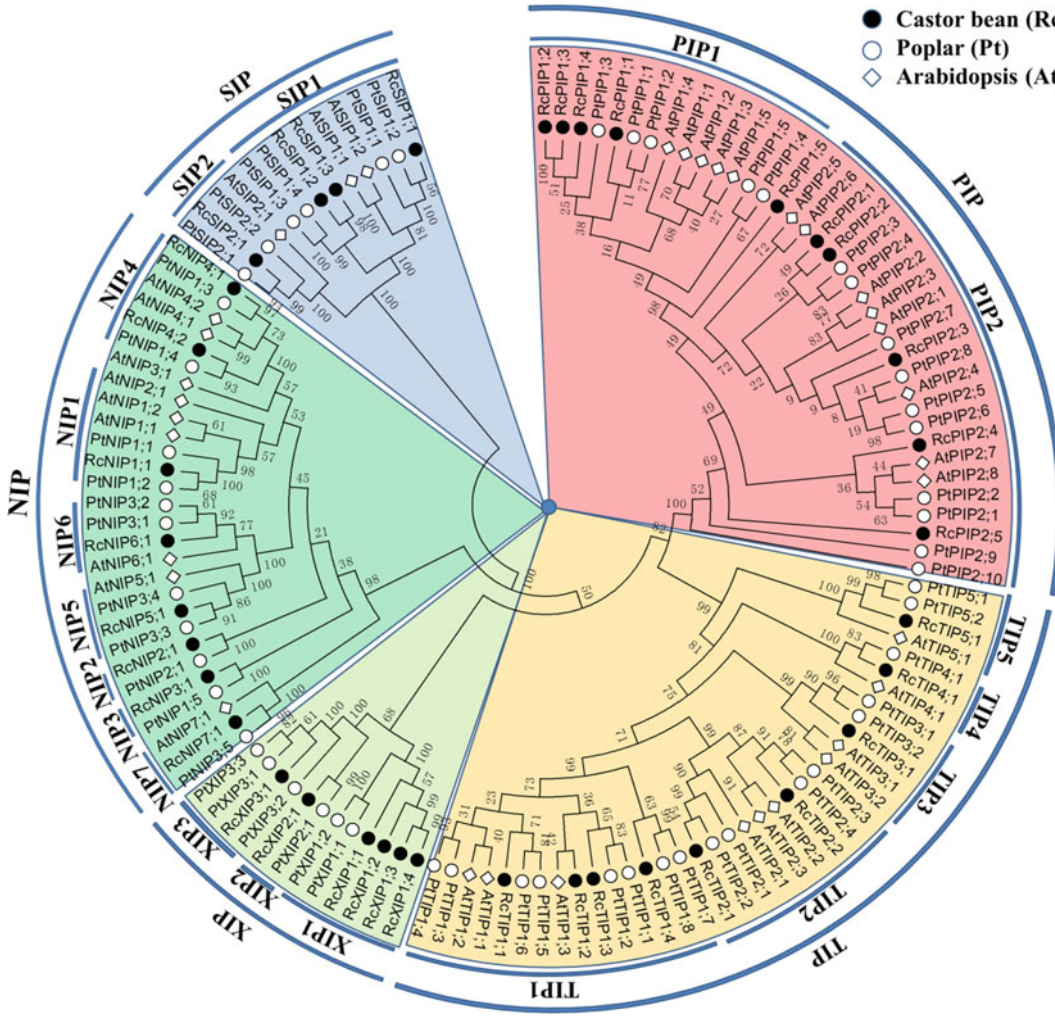

Fig. 8.2 Phylogenetic analysis of castor, Arabidopsis, and poplar AQPs (after Zou et al. 2015b)

helices was further supported by multiple alignment analysis with structure resolved *Spinacia oleracea* PIP2;1 (SoPIP2;1) and *Arabidopsis thaliana* TIP2;1 (AtTIP2;1) (Törnroth-Horsefield et al. 2006; Kirscht et al. 2016). Subcellular localization analysis indicated that RcPIPs (with average *p*I value of 7.89), RcNIPs (with average *p*I value of 8.44), RcSIPs (with average *p*I value of 9.79), and RcXIPs (with average *p*I value of 7.44) are mainly localized in plasma membranes. By contrast, RcTIPs (with average *p*I value of 5.63) are targeted to vacuoles (Zou et al. 2015b).

To learn more about the role of RcAQP genes in castor bean, RNA-seq data of leaf, flower, endosperm (II/III, V/VI), and seed were investigated. The filtered reads were mapped to the CDS of RcAQP and other protein-coding genes using Bowtie 2 (Langmead and Salzberg 2012), and the gene expression level was presented as FPKM (fragments per kilobase of exon per million fragments mapped) (Mortazavi et al. 2008). Results showed that all 37 RcAQP genes were expressed in at least one of the examined tissues (i.e., 35 in flower, 34 in leaf, 33 in seed, 31 in

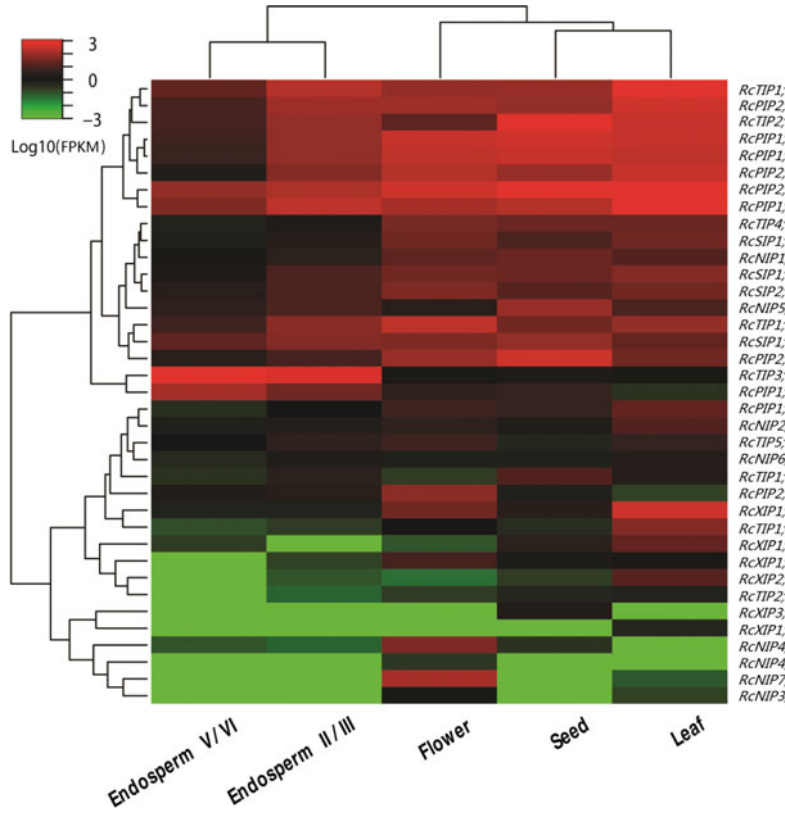

Fig. 8.3 Tissue-specific expression profiles of 37 RcAQP genes. Color scale represents fragments per kilobase of exon per million fragments mapped (FPKM) normalized $\log_{10}$ transformed counts where green indicates low expression and red indicates high expression (after Zou et al. 2015b)

stage II/III endosperm, and 29 in stage V/VI endosperm), although their transcript levels are variable. Compared with other subfamilies, expression of PIP and TIP subfamily members is considerably higher. Several key RcAQP genes were also identified in a certain tissue (Zou et al. 2015b). According to their expression patterns, RcAQP genes can be grouped into two main clusters: Cluster I includes genes that are highly expressed in most tissues, whereas Cluster II contains either genes that are expressed at low levels or genes that show tissue-specific expression (Fig. 8.3).

Oocytes of *Xenopus laevis* (African clawed frog) are commonly used for the evaluation of solute transport activity by AQPs (Preston et al. 1992). A cRNA volume of up to 50 nL can be injected in the oocyte, which allows production of large amounts of foreign proteins. Water transport activity of RcPIP1;1, RcPIP2;1, and RcTIP1;1 was tested in *X. laevis* oocytes. In these assays, RcPIP2;1 and RcTIP1;1 showed high activity, whereas RcPIP1;1 exhibited swelling rates indistinguishable from water-injected oocytes. The spatial and temporal expression of RcTIP1;1, RcPIP1;1, and RcPIP2;1 along the hypocotyl's axis was also investigated by in situ hybridization/immunolocalization as well as northern and western blots, and results showed that the amount of the RcPIP2;1 correlated best with the elongation activity of the etiolated hypocotyl and that the hydraulic conductivity of cortex cells is significantly higher in the elongating region of the hypocotyl compared with the non-elongating, mature region (Eisenbarth and Weig 2005).

8.3 PLCP Gene Family

Proteases represent dozens of unrelated families, mainly cysteine-, serine-, aspartic-, metallo-, and threonine proteases, determining the fate of all proteins (van der Hoorn 2008). In the MEROPS

database, the collected proteases were subdivided into 253 families of 61 clans based on catalytic residues as well as evolutionary relationships (Rawlings et al. 2016). Among them, clan CA, which features a nucleophilic cysteine thiol in their catalytic triad (i.e., Cys, His, and Asn), contains families of peptidases that are known to have structures similar to that of papain, a protease first discovered in papaya (*Carica papaya*) (Drenth et al. 1968; Rawlings et al. 2016). The most extensively studied papain-like cysteine proteases (PLCPs) belong to the C1A subfamily of C1 family, which also includes the C1B subfamily. Compared with C1B, C1A subfamily members contain disulfide bonds and are produced as pre-proproteases with a signal peptide, an auto-inhibitory prodomain, and a similarly sized mature protease domain (Beers et al. 2004). The signal peptide ensures that the proprotease enters the endomembrane system, whereas the prodomain prevents premature activation of the protease. Thereby, the protease precursors, mainly folding in an α-helix domain and a β-sheet domain, are usually inactive or weakly active (Turk et al. 2001). To become active, PLCPs need to be processed either by self-proteolysis or with the aid of processing enzymes, and the process depends on the pH, the action of other proteases and protease inhibitors, and the cellular or extracellular environment (van der Hoorn 2008). In animals, PLCPs are often called cathepsins, and plant PLCPs are classed as cathepsin L-, B-, H-, or F-like based on their sequence similarity to cathepsins. Furthermore, the cathepsin L-like can be divided into five phylogenetic subgroups (A–E) (Martínez and Diaz 2008). More recently, 31 PLCPs in *Arabidopsis* were classified into nine subfamilies as shown in Table 8.4 (Richau et al. 2012). Among them, the ALP subfamily members usually contain a vacuolar-targeting motif NPIR at the N terminus; the CEP subfamily members harbor a C-terminal KDEL motif for retention in the endoplasmic reticulum (ER); and most members of the RD21 and XBCP3 subfamilies carry a C-terminal extension with a proline-rich domain followed by a granulin-like domain (Ahmed et al. 2000; Hierl et al. 2012; Richau et al. 2012). PLCPs are encoded by a large gene family, and growing evidence indicated that this family is widely involved in plant growth and development, seed germination, organ senescence, immunity, and stress response (Lu et al. 2015; Misas-Villamil et al. 2016; Zou et al. 2017a, b).

The survey of the castor bean genome resulted in 26 PLCP genes, which is comparable to the 23 and 24 found in physic nut and grapevine, but relatively less than 31, 40, 40, or 43 in *Arabidopsis*, poplar, cassava, and rubber tree, respectively (Zou et al. 2017a, 2018). The RcPLCP genes are distributed across 16 scaffolds. Ten scaffolds contain a single PLCP-encoding locus, while six encode more than one: scaffold30170 (4), scaffold29646 (3), scaffold29900 (3), scaffold30131 (2), scaffold28962 (2), and scaffold29910 (2). Homology

Table 8.4 Subclassification of plant PLCP gene family

Subfamily	Full name	Name proposed by Richau et al. (2012)	Name according to Martínez and Diaz (2008)
CTB	Cathepsin B-like	CTB3-like	B-like
ALP	Aleurain-like protease	ALP-like	H-like
RD19	Responsive to dehydration 19	RD19A-like	F-like
SAG12	Senescence-associated gene 12	SAG12-like	L-like
THI	–	THI1-like	L-like
XBCP3	Xylem bark cysteine peptidase 3	XBCP3-like	L-like
XCP	Xylem cysteine peptidase	XCP2-like	L-like
CEP	Cysteine endopeptidase	CEP1-like	L-like
RD21	Responsive to dehydration 21	RD21A-like	L-like

searches showed that 13 RcPLCP genes matched EST sequences in GenBank, and the expression of other family members was supported by available RNA-seq reads.

The predicted gene structure of RcPLCP genes was manually revised. Alignment of mRNA sequences to the RcPLCP-containing scaffolds allowed the optimization of six gene models, and 10 genes (i.e., *RcRD21A*, *RcCEP1*, *RcCEP2*, *RcTHI1*, *RcSAG12H3*, *RcRD19A*, *RcRD19B*, *RcRD19C*, *RcALP1,* and *RcCTB1*) were shown to have alternative splicing isoforms (Table 8.5).

Four RcPLCP gene pairs may have originated by tandem duplications for their adjacent location on the same scaffold and their high sequence identity: 97.4% between *RcSAG12H7* (29910.t000015) and *RcSAG12H8* (29910.t000014), 96.8% between *RcSAG12H2* (28962.t000017) and *RcSAG12H3* (28962.t000018), 87.5% between *RcSAG12H4* (29646.t000033) and *RcSAG12H5* (29646.t000034), and 74.1% between *RcPAP1* (29900.t000078) and *RcPAP2* (29900.t000077). However, whether *RcPAP3* is a proximal or dispersed duplicate of *RcPAP1* and *RcPAP2* still needs to be confirmed, since the 25,878 assembled scaffolds have not been anchored to the physical chromosomes (Chan et al. 2010).

Sequence alignments between castor bean and physic nut revealed significant syntenic or nearly one-to-one orthologous relationship of PLCP genes, allowing the anchoring of 25 of the 26 RcPLCP genes to eight physic nut chromosomes (Fig. 8.4) and supporting a conserved genome evolution between these two Euphorbiaceous plants.

According to phylogenetic analysis, 26 RcPLCPs can be divided into nine subfamilies: RD21, CEP, XCP, XBCP3, THI, SAG12, RD19, ALP, and CTB, with three, two, two, two, one, eleven, three, one, or one members, respectively (Figs. 8.5 and 8.6). Most subfamilies can be further divided into several groups, which is consistent with the Best Reciprocal Hit (BRH)-based BLAST analysis: RD21 includes three groups (i.e., I, II, and III, corresponding to OG-1a, OG-1b, or OG-1c, respectively); CEP includes two groups (i.e., I and II, corresponding to OG-2a or OG-2b, respectively); XCP includes two groups (i.e., I and II, corresponding to OG-3a or OG-3b, respectively); XBCP3 includes two groups (i.e., I and II, corresponding to OG-4a or OG-4b, respectively); THI includes one group (corresponding to OG-5); SAG12 includes two groups, i.e., I (also known as PAP, corresponding to OG-6e) and II (also known as SAG12, corresponding to OG-6a, OG-6b, OG-6c, and OG-6d); RD19 includes three groups (i.e., I, II, and III, corresponding to OG-7a, OG-7b, or OG-7c, respectively); ALP includes one group (corresponding to OG-8); and CTB includes one group (corresponding to OG-9) (Table 8.5; Fig. 8.5). With its eight members, the SAG12 subfamily seems to have been highly expanded in castor bean relative to *Arabidopsis*, which contains a single SAG12 group member. On the other hand, castor bean was shown to have lost one subgroup corresponding to OG-6a, which is present in physic nut, rubber tree, cassava, and other plant species (Zou et al. 2017b). Interestingly, OG-2b, OG-4b, OG-6b, OG-6c, and OG-6d were also shown to have lost in *Arabidopsis* (Zou et al. 2018).

Analysis of exon–intron structure showed that RcPLCP genes contain one to ten introns, which is similar to what was observed in physic nut and *Arabidopsis*. Although genes in different subfamilies show distinct exon–intron structures, the structure is usually conserved within subfamily members and also between orthologs across three compared species (i.e., castor bean, physic nut, and *Arabidopsis*): All genes in the ALP subfamily contain seven introns; except for *AtSAG12* that contains two introns, other members in subfamilies SAG12 and THI feature one intron; most genes in subfamilies CEP, XCP and RD19 contain three introns, whereas *RcXCP1*, *JcXCP1*, *AtCEP1*, *AtCEP2*, *AtCEP3*, *AtRD19A,* and *AtRD19B* contain two introns instead; genes in the XBCP3 subfamily usually contain four introns, while *RcXBCP3* harbors five introns; CTB subfamily genes contain ten introns with the exception of *AtCTB1*, which has nine (Zou et al. 2018).

Table 8.5 List of 26 PLCP family genes identified in castor (after Zou et al. 2018)

Gene name	Locus ID	Transcript ID	Scaffold location	Nucleotide length (bp, from start to stop codons)		Intron no.	EST no.
				CDS	Gene		
RcRD21A	30170.t000243	30170.m013831	30170:1214722-1211316	1410	2748	4	16
RcRD21B	29801.t000069	29801.m003124	29801:415851-412454	1416	2680	4	10
RcRD21C	29970.t000002	29970.m000973	29970:17149-21809	1152	3754	3	0
RcCEP1	30147.t000097	30147.m013826	30147:2830593-2828056	1083	1636	3	3
RcCEP2	29929.t000288	29929.m004785	29929:1643874-1641686	1080	1743	3	0
RcXCP1	30162.t000046	30162.m001301	30162:1781121-1779118	1050	1262	2	0
RcXCP2	30170.t000213	30170.m013801	30170:4386238-4387922	1050	1441	3	1
RcXBCP3	30170.t000524	30170.m014112	30170:2901979-2898207	1401	3263	5	1
RcXBCP3L	29381.t000001	29381.m000072	29381:7678-12160	1506	4315	4	13
RcTHI1	29646.t000057	29646.m001109	29646:343762-345957	1044	1534	1	1
RcSAG12H1	30131.t000408	30131.m007257	30131:2504200-2506430	1089	1169	1	0
RcSAG12H2	28962.t000017	28962.m000448	28962:92944-94101	1023	1111	1	0
RcSAG12H3	28962.t000018	28962.m000449	28962:96181-97735	1023	1111	1	0
RcSAG12H4	29646.t000033	29646.m001085	29646:207526-209015	1050	1228	1	0
RcSAG12H5	29646.t000034	29646.m001086	29646:211496-213036	1029	1221	1	0
RcSAG12H6	29900.t000065	29900.m001603	29900:407069-405639	1035	1144	1	0
RcSAG12H7	29910.t000015	29910.m000924	29910:208876-206709	1026	1908	1	0
RcSAG12H8	29910.t000014	29910.m000923	29910:204533-202640	1029	1894	1	0
RcPAP1	29900.t000078	29900.m001616	29900:487718-489114	1023	1124	1	0
RcPAP2	29900.t000077	29900.m001615	29900:483819-485847	1032	1812	1	1
RcPAP3	29827.t000145	29827.m002672	29827:836744-835443	1029	1280	1	0
RcRD19A	30131.t000249	30131.m007098	30131:1514766-1516903	1122	1683	3	8
RcRD19B	30170.t000534	30170.m014122	30170:2964578-2966528	1101	1547	3	68
RcRD19C	28462.t000004	28462.m000130	28462:46261-49312	1146	2404	3	17
RcALP1	29739.t000193	29739.m003757	29739:1197581-1200666	1077	2458	7	3
RcCTB1	30076.t000074	30076.m004510	30076:399629-403162	1080	2789	10	2

(continued)

Table 8.5 (conitnued)

Gene name	Alternative splicing	Comment	Jc_ortholog	At_ortholog	Subfamily (group)	Orthologous group
RcRD21A	Yes	–	JcRD21A	AtRD21A	RD21 (I)	OG-1a
RcRD21B	–	–	JcRD21B	–	RD21 (II)	OG-1b
RcRD21C	–	Misannotated	JcRD21C	AtRDL1	RD21 (III)	OG-1c
RcCEP1	Yes	–	JcCEP1	AtCEP1	CEP (I)	OG-2a
RcCEP2	Yes	–	JcCEP2	–	CEP (II)	OG-2b
RcXCP1	–	Misassembled	JcXCP1	AtXCP1	XCP (I)	OG-3a
RcXCP2	–	–	JcXCP2	AtXCP2	XCP (II)	OG-3b
RcXBCP3	–	Misannotated	JcXBCP3	AtXBCP3	XBCP3 (I)	OG-4a
RcXBCP3L	–	Misannotated	JcXBCP3L	–	XBCP3 (II)	OG-4b
RcTHI1	Yes	–	JcTHI1	AtTHI1	THI	OG-5
RcSAG12H1	–	–	JcSAG12H3	AtSAG12	SAG12 (I)	OG-6b
RcSAG12H2	–	–	JcSAG12H4	AtSAG12	SAG12 (I)	OG-6b
RcSAG12H3	Yes	–	JcSAG12H4	AtSAG12	SAG12 (I)	OG-6b
RcSAG12H4	–	–	–	AtSAG12	SAG12 (I)	OG-6d
RcSAG12H5	–	–	–	AtSAG12	SAG12 (I)	OG-6d
RcSAG12H6	–	–	JcSAG12H7	AtSAG12	SAG12 (I)	OG-6c
RcSAG12H7	–	–	JcSAG12H8	AtSAG12	SAG12 (I)	OG-6c
RcSAG12H8	–	–	JcSAG12H8	AtSAG12	SAG12 (I)	OG-6c
RcPAP1	–	–	–	AtPAP1	SAG12 (II)	OG-6e
RcPAP2	–	–	–	AtPAP1	SAG12 (II)	OG-6e
RcPAP3	–	Misannotated	–	AtPAP1	SAG12 (II)	OG-6e
RcRD19A	Yes	–	JcRD19A	AtRD19A	RD19 (I)	OG-7a
RcRD19B	Yes	–	JcRD19B	AtRD19C	RD19 (II)	OG-7b
RcRD19C	Yes	–	JcRD19C	AtRD19D	RD19 (III)	OG-7c
RcALP1	Yes	–	JcALP1	AtALP	ALP	OG-8
RcCTB1	Yes	Misannotated	JcCTB1	AtCTB2	CTB	OG-9

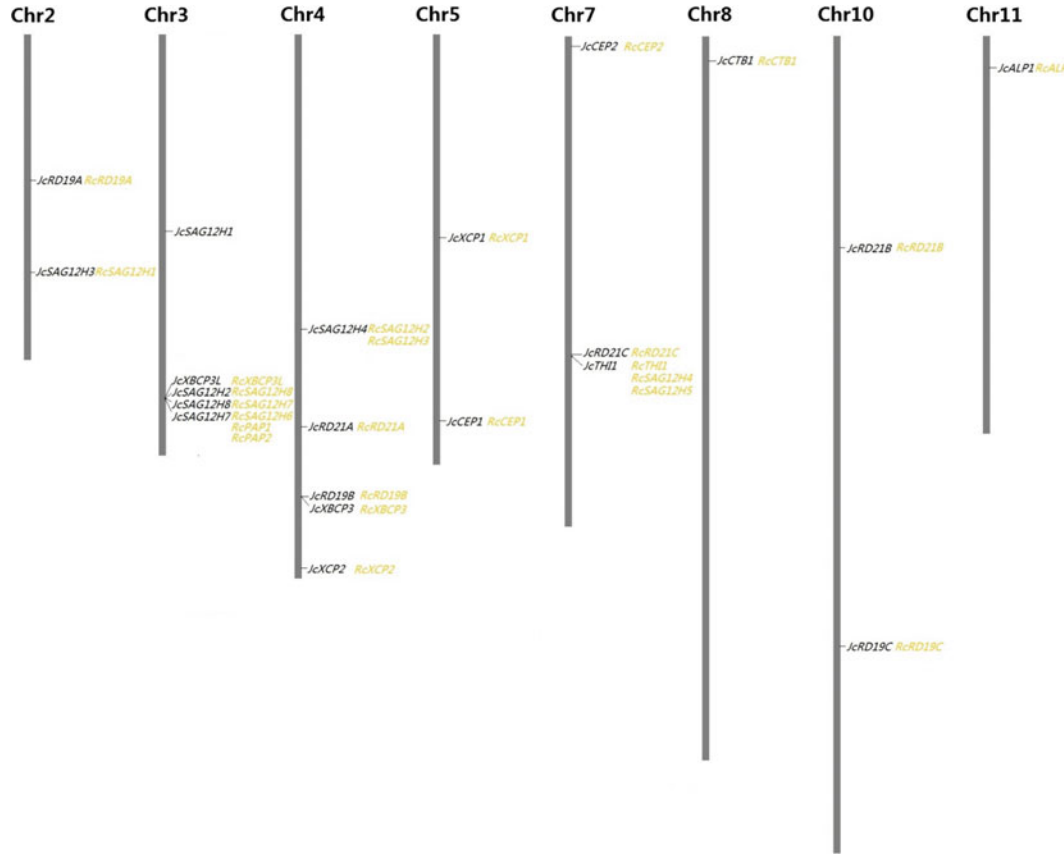

Fig. 8.4 Anchoring RcPLCP genes to eight physic nut chromosomes. The chromosome serial number is indicated at the top of each chromosome, and 25 RcPLCP genes were shown just behind their collinear genes in physic nut (where *RcSAG12H4*, *RcSAG12H5*, *RcPAP1,* and *RcPAP2* with no collinear genes in physic nut) (after Zou et al. 2018)

Transcriptional profiling revealed that 26 RcPLCP genes were expressed in at least one of the examined tissues or developmental stages (i.e., 20 in leaf, 23 in male flower, 19 in endosperm II/III, 16 in endosperm V/VI, 21 in developing seed, and 17 in germinating seed). Though at variable levels, most RcPLCP genes showed expression in all examined tissues. *RcCEP1* represents the most highly expressed PLCP gene in endosperm II/III, endosperm V/VI, developing seed, and germinating seed, accounting for about 46, 41, 43, or 27% of the total PLCP transcripts in each sample, respectively. By contrast, its expression level in leaf was very low. *RcCEP2* showed comparable expression level to that of *RcCEP1* in endosperm V/VI, but its expression level is lower in other tissues. *RcTHI1* represents the most expressed gene among PLCP family members in flower, while *RcRD19B* and *RcRD21A* represent two of the most expressed genes in leaf, accounting for about 29 or 25% of the total PLCP transcripts, respectively. Nevertheless, the expression of *RcRD19B* was not detected in endosperm V/VI. *RcSAG12H5*, *RcSAG12H6*, and *RcPAP1* seem to be flower-specific, and *RcPAP3* seems to be leaf-specific (Fig. 8.7).

According to the expression pattern over various tissues, 26 RcPLCP genes can be grouped into four main clusters: Cluster I is predominantly expressed in germinating seed, and it includes *RcCEP1, RcCEP2*, *RcRD21B, RcRD21A*, *RcRD 19A*, *RcALP1*, *RcCTB1,* and *RcXBCP3*. Cluster II is preferentially expressed in developing seed and

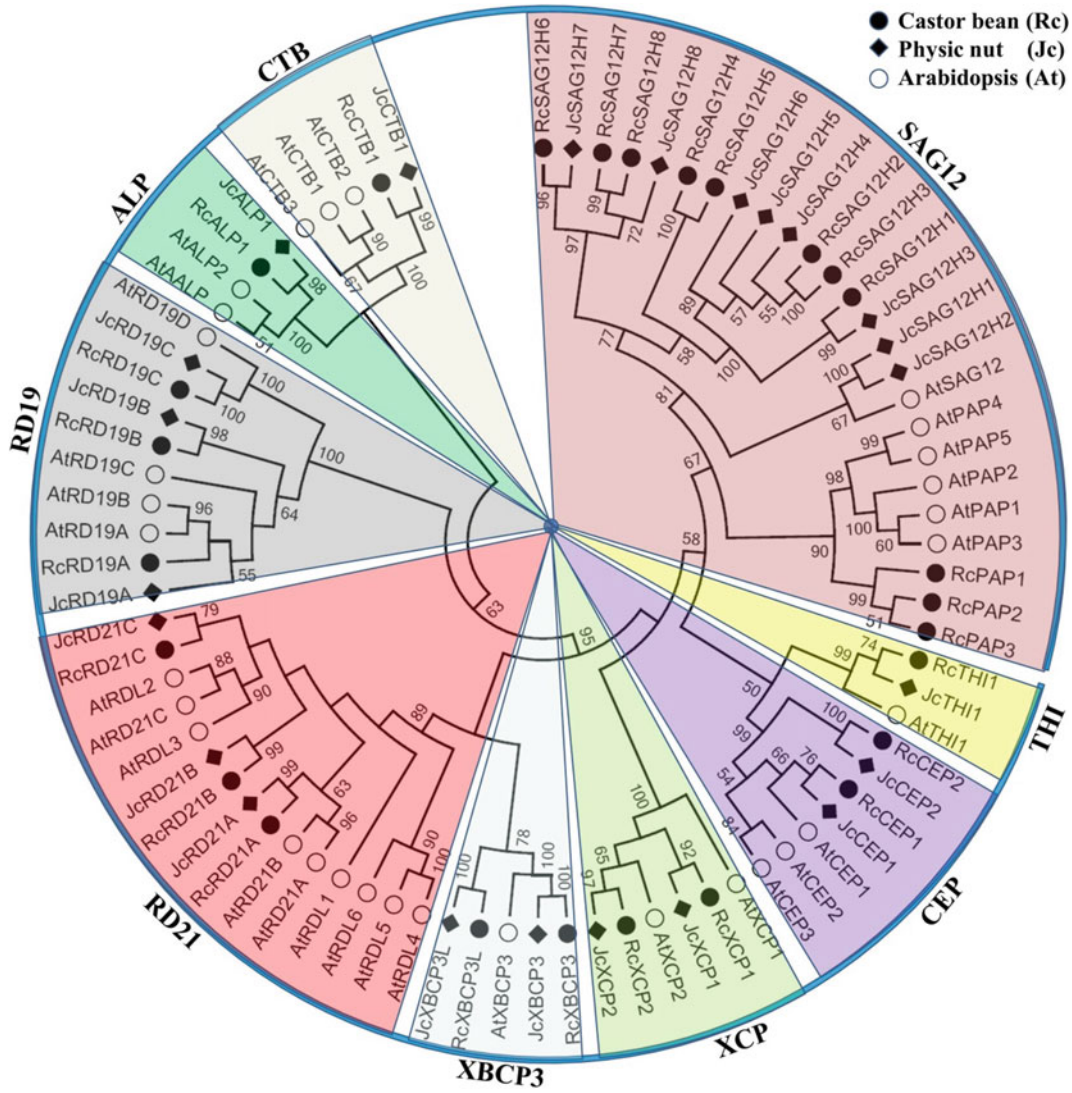

Fig. 8.5 Phylogenetic analysis of castor bean, physic nut, and *Arabidopsis* PLCPs (after Zou et al. 2018)

endosperm, and it is composed of *RcRD19C*, *RcSAG12H7*, and *RcSAG12H8*. Cluster III is mostly expressed in male flower, and it includes *RcTHI1*, *RcSAG12H1*, *RcSAG12H2*, *RcSAG 12H3*, *RcSAG12H4*, *RcSAG12H5*, *RcSAG12H6*, and *RcPAP1*. Cluster IV is typically expressed in leaf and includes *RcXCP1*, *RcXCP2*, *RcPAP2, RcPAP3*, *RcRD21C*, *RcXBCP3L,* and *RcRD19B* (Fig. 8.7).

Moreover, RcCEP1 (also known as CysEP or CysEP1) has been shown to be involved in programmed cell death of castor endosperm (Schmid et al. 1998, 1999). RcCEP1 is synthesized as a pre-proenzyme and the pre-sequence is removed in the endoplasmic reticulum (ER). In germinating seeds, the 45 kDa proenzyme is transported from the ER to the cytosol of senescing endosperm cells within ER-derived vesicles called ricinosomes. Disruption of vacuoles and subsequent acidification of ricinosomes cause activation of RcCEP1, with cleavage of the N-terminal propeptide and the C-terminal KDEL motif. The 35 kDa mature protease degrades cell proteins such as

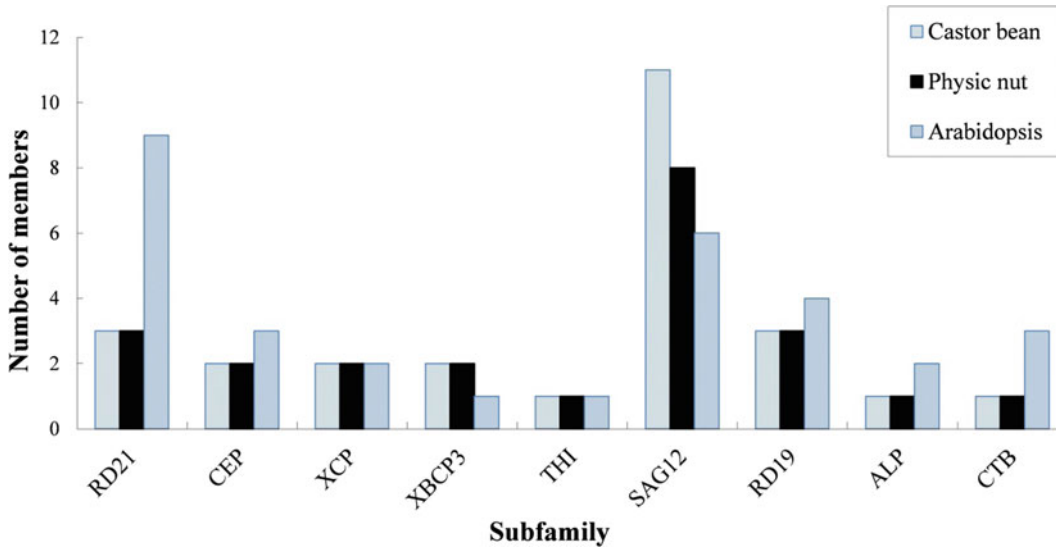

Fig. 8.6 Distribution of castor bean, physic nut, and *Arabidopsis* PLCP family genes in subfamilies (after Zou et al. 2018)

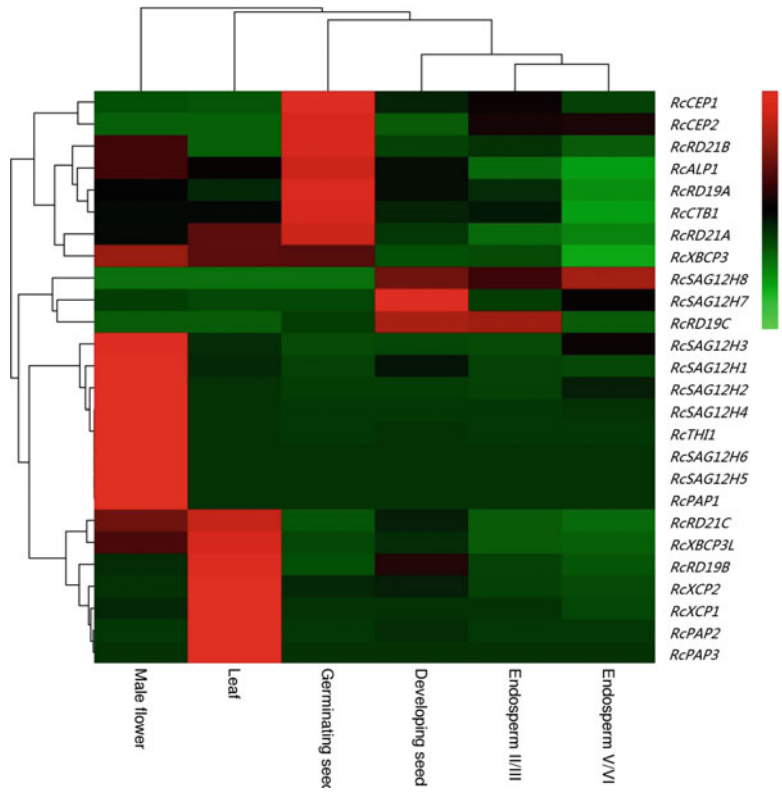

Fig. 8.7 Tissue-specific expression profiles of 26 RcPLCP genes. Color scale represents FPKM normalized $\log_{10}$ transformed counts where green indicates low expression and red indicates high expression (after Zou et al. 2018)

glyoxysomal malate dehydrogenase precursor and hydroxyproline (Hyp)-rich proteins (extensins) that form the basic scaffold of the plant cell wall (Schmid et al. 2001; Than et al. 2004).

8.4 Lhc Supergene Family

Oxygenic photosynthesis, which converts light energy into chemical energy, takes place in photosystems I and II (PSI/II) of green plants, algae as well as cyanobacteria. Each photosystem is composed of a core complex and a peripheral antennae system (i.e., LHCI/II, light-harvesting complex of PSI/II) adjacent to the reaction center. Light is captured by pigments such as chlorophylls and carotenoids that bind to antennae proteins (Croce and van Amerongen 2014; Rochaix 2014). Nuclear-encoded antennae proteins constitute the so-called light-harvesting chlorophyll a/b-binding (Lhc) superfamily, which was characterized by the presence of a chlorophyll-binding domain in the transmembrane alpha-helix (Jansson 1999). Several distinct gene families have been described in this superfamily, including chlorophyll a/b-binding protein (CAB) or Lhc, photosystem II subunit S (PsbS), light-harvesting-like protein (Lil), ferrochelatase II (FCII), and red lineage chlorophyll a/b-binding-like protein (RedCAP) (Jansson 1999; Klimmek et al. 2006; Engelken et al. 2010). Among them, RedCAP, as the name suggests, is confined to the red lineage (Engelken et al. 2010). Lhc family proteins, featuring three transmembrane alpha-helixes and mainly functioning in light absorption and energy transfer to reaction centers, form two evolutionary subfamilies (i.e., Lhca and Lhcb) that are named for their association with PSI or PSII, respectively. By contrast, other families were shown to be involved in stress protection (Jansson 1999; Klimmek et al. 2006). In contrast to four alpha-helixes present in PsbS proteins, one to three helixes can be found in Lil proteins, which is dependent on different subfamilies: One-helix proteins (OHPs) include two distinct groups, i.e., OHP1/Lil2 and OHP2/Lil6; two helixes are present in stress-enhanced protein (SEP), which includes SEP1/Lil4, SEP2/Lil5, SEP3/Lil3, SEP4, and SEP5; three helixes are present in early light-induced protein (ELIP) or Lil1; a partial chlorophyll-binding domain is present in photosystem II protein 33 (Psb33) or Lil8. Additionally, FCII also contains one chlorophyll-binding domain as well as one ferrochelatase domain in higher plants (Engelken et al. 2010). Tracing their origin suggested that OHP is more primitive, which is more likely to be derived from the plastid-encoded high light-induced protein (HLIP) via gene transfer after the primary endosymbiosis (Koziol et al. 2007; Engelken et al. 2010).

The search of the castor bean genome resulted in 28 Lhc superfamily genes from 26 scaffolds, and the gene number is comparable to 27 in physic nut but relatively less than 34 in *Arabidopsis* (Table 8.6). Although most scaffolds encode a single Lhc superfamily gene, two of them contain two, i.e., scaffold30005 (*RcLhca3* and *RcLhcb1.2*) and scaffold30170 (*RcLhca5* and *RcLhcb7*). The expression of these genes was all supported by ESTs and/or RNA-seq reads, which allowed the optimization of two gene models, i.e., *RcSEP1* and *RcSEP5* (Table 8.6; Fig. 8.8).

Based on optimized gene models, the intron number of RcLhc superfamily genes was shown to vary from zero to nine. The average length of CDS is about 765 bp, ranging from 348 to 1533 bp. By contrast, the gene length (from start to stop codons) is more variable, with the average of 1694 bp and the ranging from 576 to 7944 bp (Fig. 8.8). The deduced proteins were shown to consist of 115–510 amino acids, and the predicted length of transit peptide (TP) varies from 30 to 97 residues (Table 8.6).

According to BRH-based homologous analysis, the majority of RcLhc superfamily genes were shown to have orthologs in physic nut and *Arabidopsis*, and nearly one-to-one orthologous relationship was observed between castor bean

Table 8.6 List of 28 Lhc superfamily genes identified in castor (after Zou et al. 2013)

Gene name	Locus ID	Transcript ID	Position	Protein length	TP length	Group and comments	Jc_ortholog	At_ortholog	Orthologous group
RcLhca1	28802.t000006	28802.m000133	28802:64842-66230	244	44	Lhca	JcLhca1	AtLhca1	Lhca1
RcLhca2	29929.t000153	29929.m004650	29929:901616-899794	270	58	Lhca	JcLhca2	AtLhca2	Lhca2
RcLhca3	30005.t000015	30005.m001249	30005:75318-73998	273	40	Lhca	JcLhca3	AtLhca3	Lhca3
RcLhca4	30138.t000234	30138.m004060	30138:1605408-1606729	251	50	Lhca	JcLhca4	AtLhca4	Lhca4
RcLhca5	30170.t000292	30170.m013880	30170:1526617-1528160	269	47	Lhca	JcLhca5	AtLhca5	Lhca5
RcLhca6	29739.t000165	29739.m003729	29739:1053566-1055670	297	76	Lhca	JcLhca6	AtLhca6	Lhca6
RcLhcb1.1 *RcLhcb1.2* *RcLhcb1.3*	29693.t000089 30005.t000066 30076.t000142	29693.m002051 30005.m001300 30076.m004578	29693:622387-623389 30005:435512-436703 30076:908573-909570	265 265 267	35 35 36	Lhcb	JcLhcb1.1 JcLhcb1.2	AtLhcb1.1 AtLhcb1.2 AtLhcb1.3 AtLhcb1.4 AtLhcb1.5	Lhcb1
RcLhcb2	29597.t000004	29597.m000282	29597:35750-37662	265	37	Lhcb	JcLhcb2	AtLhcb2.1 AtLhcb2.2 AtLhcb2.3	Lhcb2
RcLhcb3	30027.t000008	30027.m000815	30027:110457-109075	268	45	Lhcb	JcLhcb3	AtLhcb3	Lhcb3
RcLhcb4	29701.t000022	29701.m000601	29701:179105-180768	285	30	Lhcb	JcLhcb4	AtLhcb4.1 AtLhcb4.2	Lhcb4
RcLhcb5	29827.t000140	29827.m002667	29827:809734-811736	292	49	Lhcb	JcLhcb5	AtLhcb5	Lhcb5
RcLhcb6	28509.t000004	28509.m000094	28509:57886-59242	256	47	Lhcb	JcLhcb6	AtLhcb6	Lhcb6
RcLhcb7	30170.t000221	30170.m013809	30170:4506974-4509648	329	43	Lhcb	JcLhcb7	AtLhcb7	Lhcb7
RcLhcb8	29669.t000052	29669.m000843	29669:328491-326418	282	32	Lhcb	JcLhcb8	AtLhcb8	Lhcb8
RcPsbS	29908.t000190	29908.m006139	29908:145579-147478	275	70	PsbS	JcPsbS	AtPsbS	PsbS
RcELIP	29844.t000066	29844.m003220	29844:442084-443101	191	48	ELIP	JcELIP	AtELIP1 AtELIP2	ELIP
RcOHP1	29568.t000003	29568.m000280	29568:46263-46853	115	45	OHP	JcOHP1	AtOHP1	OHP1
RcOHP2	29863.t000044	29863.m001084	29863:304660-302446	185	49	OHP	JcOHP2	AtOHP2	OHP2
RcSEP1	29801.t000053	29801.m003108	29801:318811-322414	143	71	SEP, misannotated	JcSEP1	AtSEP1	SEP1
RcSEP2	29719.t000019	29719.m000332	29719:233637-235422	201	32	SEP	JcSEP2	AtSEP2	SEP2
RcSEP3	29738.t000005	29738.m001003	29738:44102-41691	259	56	SEP	JcSEP3	AtSEP3.1 AtSEP3.2	SEP3

(continued)

Table 8.6 (continued)

Gene name	Locus ID	Transcript ID	Position	Protein length	TP length	Group and comments	Jc_ortholog	At_ortholog	Orthologous group
RcSEP4	29758.t000037	29758.m000674	29758:340934-339994	191	39	SEP	JcSEP4	AtSEP4	SEP4
RcSEP5	28456.t000001	28456.m000031	28456:10790-6820	144	54	SEP, misannotated	JcSEP5	AtSEP5	SEP5
RcSEP6	30147.t000625	30147.m014354	30147:1502524-1501557	240	69	SEP	JcSEP6	–	SEP6
RcPsb33	29938.t000027	29938.m000610	29938:159986-162387	280	57	Psb33	JcPsb33	AtPsb33	Psb33
RcFCII	30225.t000055	30225.m001731	30225:2273280-2281223	510	97	FCII	JcFCII	AtFCII	FCII

and physic nut. By contrast, the orthologous relationship between castor bean and *Arabidopsis* was shown to be relatively complex, including one-to-one, one-to-two, one-to-three, and two-to-five. The result was further supported by synteny analysis: Among ten duplicates found in *Arabidopsis*, four are more likely to be derived from tandem duplication, five from the α WGD, and one from the γ WGD (i.e., *AtLhcb8*); in physic nut, one duplicate may be resulted from tandem duplication and two from the γ WGD (i.e., *JcLhcb8* and *JcSEP6*); in castor bean, one duplicate may originate by dispersed duplication and three from the γ WGD (i.e., *RcLhcb1.3*, *JcLhcb8,* and *JcSEP6*). It is worth mentioning that no ortholog was found for *RcSEP6* in *Arabidopsis*, and for *RcLhcb1.3* in both physic nut and *Arabidopsis*. *RcLhcb1.2*, which may originate by dispersed duplication, is located on scaffold30005 together with *RcLhca3*. However, no collinear gene was found for *RcLhcb1.2* in physic nut (Fig. 8.9). Based on these results, 28 RcLhc superfamily genes were assigned to 26 orthologous groups (OGs) representing four gene families (i.e., *Lhc*, *PsbS*, *Lil,* and *FCII*) (Table 8.6), and these OGs served as a basis for gene expansion in core eudicots and other lineages.

According to transcriptional profiling, RcLhc superfamily genes were shown to be expressed in at least one of five examined tissues (i.e., leaf, male flower, endosperm II/III, endosperm V/VI, and germinating seed), though their transcript levels are highly variable. Except for *RcSEP6* that is predominantly expressed in germinating seed, other genes are preferentially expressed in leaf, where the total transcripts were about 64-, 90-, 198-, or 365-folds than that in male flower, germinating seed, endosperm II/III, and endosperm V/VI, respectively. Several key members were also identified in a certain tissue: *RcLhcb1.2* represents the most highly expressed gene among RcLhc superfamily members in leaf and male flower; *RcSEP6* represents the most expressed gene in germinating seed; *RcSEP3* represents the most expressed gene in endosperm II/III and endosperm V/VI (Fig. 8.10).

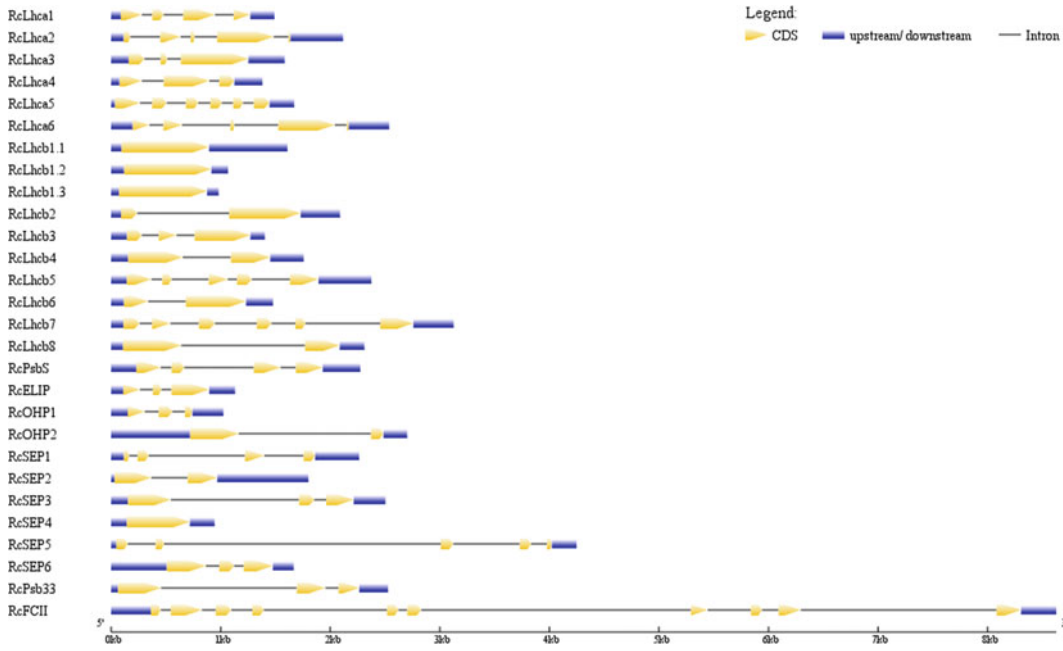

Fig. 8.8 Exon–intron structures of 28 RcLhc superfamily genes. Graphic representation of the gene structures was displayed using GSDS (http://gsds.cbi.pku.edu.cn/)

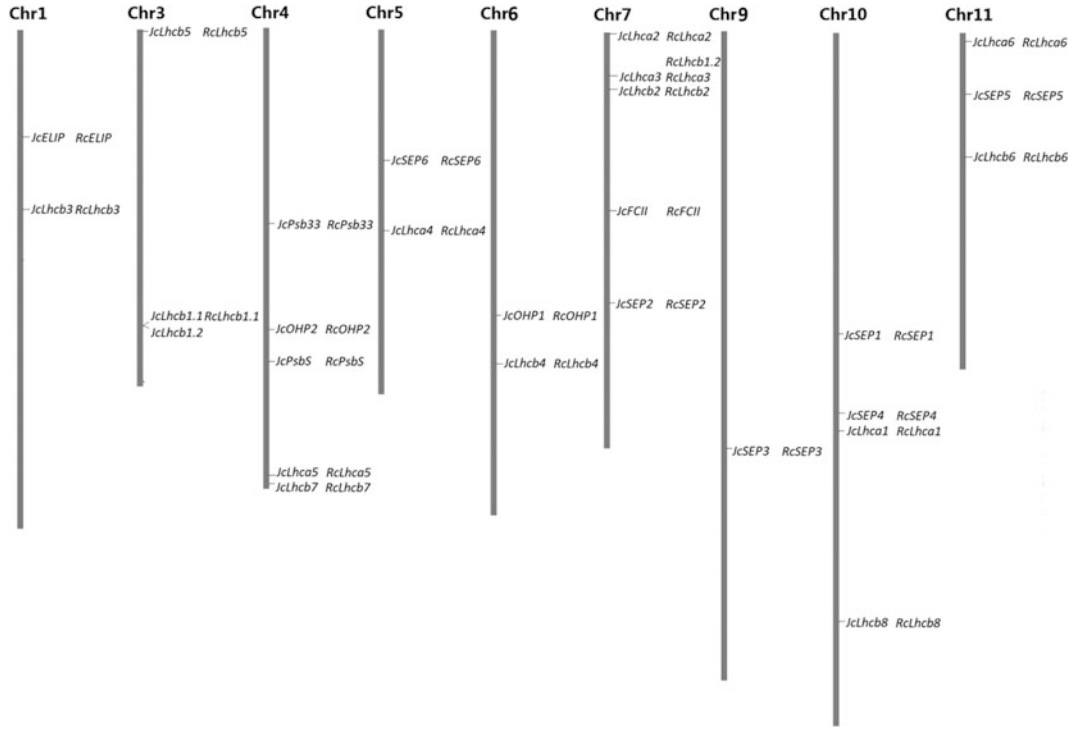

Fig. 8.9 Anchoring RcLhc superfamily genes to nine physic nut chromosomes. The chromosome serial number is indicated at the top of each chromosome, and 27 RcLhc superfamily genes were shown just behind their collinear genes in physic nut (where *RcLhcb1.2* with no collinear gene in physic nut)

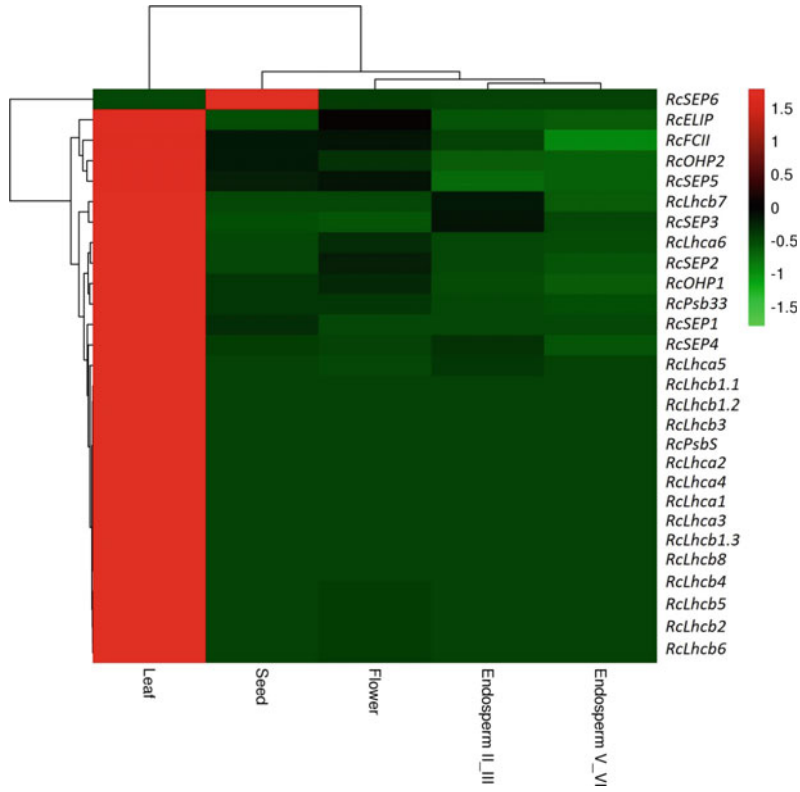

Fig. 8.10 Tissue-specific expression profiles of 28 JcLhc superfamily genes. Color scale represents FPKM normalized $\log_{10}$ transformed counts where green indicates low expression and red indicates high expression (after Zou et al. 2013)

8.5 WRKY Transcription Factor Gene Family

WRKY transcription factors, defined by the presence of the conserved WRKY domain of approximate 60 amino acids, play an essential regulatory role in plant growth, development, metabolism, and biotic and abiotic stress responses (Rushton et al. 2010; Chen et al. 2012; Schluttenhofer and Yuan 2015). Since the first WRKY gene was isolated from sweet potato (*Ipomoea batatas*), homologs have been identified in a wide range of plants and several non-plant species, including some diplomonads, social amoebae and other amoebozoa, and fungi such as *incertae sedis*. Genome-wide surveys revealed that the WRKY gene family in vascular plants has been shown to be highly expanded and diversified relative to non-vascular plants or other kingdoms (Zhang and Wang 2005; Rinerson et al. 2015). WRKY proteins contain one or two WRKY domains, comprising the highly conserved WRKYGQK heptapeptide at the N-termini and a zinc finger motif ($CX_{4-7}CX_{22-23}HXH/C$) at C-termini. These two motifs are vital for high-affinity binding to a cis-acting element termed W box (TTGACT/C) (Eulgem et al. 2000) which is present in the regulatory regions of target genes. According to the number of WRKY domains and the features of their zinc finger motifs, WRKY proteins can be categorized into three main groups. Group I members contain two WRKY domains and feature the zinc finger motif of C_2H_2. Group II contains a single WRKY domain and a C_2H_2 zinc finger, and Group III possesses a single WRKY domain and a C_2HC zinc finger motif. Based on the presence of certain amino acid motifs outside the WRKY domain, Group II can be further divided into five subgroups (IIa to IIe), where distinct intron location was found among different subgroups (Eulgem et al. 2000). Subgroups IIc, IId, and IIe contain a conserved PR intron located after the

codon encoding arginine of the N-terminal to the zinc finger motif, which is also found Group III and the C-terminal WRKY domain of Group I. By contrast, subgroups IIa and IIb feature a VQR intron in the zinc finger motif (Eulgem et al. 2000; Rinerson et al. 2015).

A search of the castor bean genome identified 58 WRKY family genes that are distributed across 41 scaffolds (Zou et al. 2016b), and the family number is comparable to 58 found in physic nut and 59 described in grapevine (Xiong et al. 2013; Guo et al. 2014). *Arabidopsis* and poplar, on the other hand, showed a higher number of family members (72 and 105, respectively) (Eulgem et al. 2000; He et al. 2012). The expression of RcWRKY genes was all supported by available mRNA sequences, which allowed optimization of 20 out of the 57 predicted WRKY gene models. Moreover, one additional locus encoding 117 residues was identified in scaffold28842 (Table 8.7).

Eleven RcWRKY genes belong to Group I as they contain the N-terminal motif $Cx_4Cx_{22–23}HxH$ and C-terminal $Cx_4Cx_{23}HxH$. Thirty-nine were classified as Group II with a zinc finger motif of the typical $Cx_{4–5}Cx_{23}HxH$, and eight correspond to Group III with the motif $Cx_7Cx_{23}HxC$. RcWRKY genes in Group II could be further divided into subgroup IIa, IIb, IIc, IId, and IIe, with 3, 8, 16, 6, and 6 members, respectively (Figs. 8.11 and 8.12).

A significant level of syntenic relation and a nearly one-to-one orthologous relationship were observed between castor bean and physic nut WRKY genes. Albeit, castor bean contains one more Group III member that was probably derived from proximal duplication (i.e. *RcWRKY55* and *RcWRKY56*, separated by 39 genes), and lacks an ortholog of *JcWRKY25*. Moreover, *RcWRKY29* does not have the second intron observed in *JcWRKY28* (Zou et al. 2016b).

Most RcWRKY proteins were shown to harbor the conserved heptapeptide WRKYGQK (Fig. 8.11). However, three family members (i.e. RcWRKY12, RcWRKY13, and RcWRKY14) showed the WRKYGKK variety as seen in physic nut, *Arabidopsis*, and other plant species. A conserved leucine zipper motif, which is involved in protein dimerization and DNA binding, was found in RcWRKY29 and RcWRKY33. A HARF motif was found in three subgroup IId members (i.e. RcWRKY39, RcWRKY41, and RcWRKY43). The active repressor motif LxLxLx was found in two subgroup IIa members (i.e. RcWRKY28 and RcWRKY30) and four subgroup IIb members (i.e. RcWRKY35, RcWRKY36, RcWRKY37, and RcWRKY38). Additionally, subgroup-specific motifs were also identified by using MEME (Fig. 8.13).

The expression profiles of RcWRKY genes were analyzed by RNA-seq and quantitative reverse transcriptase PCR (qRT-PCR) (Li et al. 2012a; Zou et al. 2016b). Expression of all RcWRKY genes was detected in at least one of the examined tissues (55 in leaf, 51 in male flower, 51 in endosperm, and 51 in seed). According to the FPKM analysis, total transcripts of the whole family were relatively lower in endosperm than that in the other three tissues analyzed, and the total transcripts in stage II/III endosperm were twofold higher than that in stage V/VI endosperm (Fig. 8.14). Among the 24 genes that showed differential expression between the two endosperm developmental stages, 23 were down-regulated and only one was up-regulated in stage V/VI endosperm, suggesting regulatory roles in early endosperm development.

In vivo experiment showed that endogenous abscisic acid (ABA) levels are closely associated with storage material accumulation in developing castor bean seeds (Chandrasekaran and Liu 2014). In vitro, exogenous ABA also enhanced the dry weight (including the accumulation of soluble sugar and total lipid content) of developing seeds cultured in a nutrient medium (Chandrasekaran et al. 2014). After the application of 10 μM ABA for 24 h, differential gene expression analysis indicated that 2568 genes were up- or down-regulated at least two folds (Chandrasekaran et al. 2014), which include 13 out of the 58 *RcWRKY* genes (Fig. 8.15). Among them, eleven (four Group I members, two

Table 8.7 WRKY family genes in castor bean (after Zou et al. 2016b)

Gene	Locus ID	Scaffold	Identified position	EST hits	AS[a]	AS[b]	(Sub)group and comments	At_ortholog	Jc_ortholog
RcWRKY01	29949.t000007	scaffold29949	49111-52359	–	–	Yes	I	AtWRKY01	JcWRKY01
RcWRKY02	27613.t000032	scaffold27613	207484-214016	2	–	Yes	I	AtWRKY20	JcWRKY09
RcWRKY03	29929.t000090	scaffold29929	509685-512932	–	–	–	I, misassembled	AtWRKY20	JcWRKY10
RcWRKY04	28966.t000003	scaffold28966	29398-23813	–	–	Yes	I	AtWRKY02	JcWRKY11
RcWRKY05	29717.t000002	scaffold29717	13200-10572	43	Yes	Yes	I	AtWRKY33	JcWRKY07
RcWRKY06	29820.t000050	scaffold29820	294438-289616	–	–	Yes	I, misannotated	AtWRKY33	JcWRKY08
RcWRKY07	29635.t000028	scaffold29635	203640-198543	1	–	Yes	I, misannotated	AtWRKY04	JcWRKY06
RcWRKY08	30174.t000563	scaffold30174	3380374-3384508	9	–	–	I, misannotated	AtWRKY04	JcWRKY05
RcWRKY09	29805.t000035	scaffold29805	187614-192328	–	–	Yes	I	AtWRKY44	JcWRKY04
RcWRKY10	29687.t000003	scaffold29687	15455-21624	1	–	Yes	I, misannotated	AtWRKY32	JcWRKY02
RcWRKY11	29848.t000095	scaffold29848	494292-491256	–	–	Yes	I, misannotated	AtWRKY32	JcWRKY03
RcWRKY12	29771.t000001	scaffold29771	4111-8868	–	–	–	IIc	AtWRKY51	JcWRKY12
RcWRKY13	29739.t000022	scaffold29739	137722-136462	–	–	Yes	IIc	AtWRKY50	JcWRKY14
RcWRKY14	28644.t000022	scaffold28644	112007-114711	–	–	–	IIc	AtWRKY51	JcWRKY13
RcWRKY15	29929.t000127	scaffold29929	718359-720760	–	–	–	IIc	AtWRKY75	JcWRKY17
RcWRKY16	30190.t000144	scaffold30190	611849-613809	–	–	–	IIc	AtWRKY75	JcWRKY18
RcWRKY17	30147.t000745	scaffold30147	2203665-2204574	–	–	–	IIc	AtWRKY75	JcWRKY19
RcWRKY18	30174.t000066	scaffold30174	2075781-2078009	–	–	Yes	IIc, misannotated	AtWRKY43	JcWRKY21
RcWRKY19	30190.t000514	scaffold30190	3026363-3027328	–	–	–	IIc	AtWRKY56	JcWRKY20
RcWRKY20	28040.t000001	scaffold28040	32051-27986	–	–	–	IIc, misannotated	AtWRKY13	JcWRKY15
RcWRKY21	29709.t000007	scaffold29709	41208-44408	–	–	Yes	IIc, misannotated	AtWRKY12	JcWRKY16
RcWRKY22	29889.t000087	scaffold29889	412309-414133	1	–	–	IIc, misannotated	AtWRKY48	JcWRKY23
RcWRKY23	29693.t000098	scaffold29693	677801-676041	–	–	–	IIc	AtWRKY28	JcWRKY26
RcWRKY24	30076.t000187	scaffold30076	1245777-1243970	–	–	–	IIc	AtWRKY23	JcWRKY22
RcWRKY25	30174.t000532	scaffold30174	3225599-3218630	1	–	Yes	IIc, misannotated	AtWRKY57	JcWRKY24
RcWRKY26	29767.t000010	scaffold29767	166564-163614	–	–	–	IIc	AtWRKY49	JcWRKY38

(continued)

Table 8.7 (continued)

Gene	Locus ID	Scaffold	Identified position	EST hits	AS[a]	AS[b]	(Sub)group and comments	At_ortholog	Jc_ortholog
RcWRKY27	N/A	scaffold28842	266259-272883	–	–	Yes	IIc, not predicted	–	JcWRKY58
RcWRKY28	30131.t000001	scaffold30131	4396-6722	7	–	Yes	IIa, misassembled	AtWRKY40	JcWRKY27
	43951.t000001	scaffold43951							
RcWRKY29	29848.t000101	scaffold29848	538811-533980	1	–	Yes	IIa, misannotated	AtWRKY40	JcWRKY28
RcWRKY30	29848.t000100	scaffold29848	523248-521822	–	–	Yes	IIa, misannotated	AtWRKY40	JcWRKY29
RcWRKY31	29842.t000052	scaffold29842	272719-275658	1	Yes	Yes	IIb	AtWRKY42	JcWRKY31
RcWRKY32	30010.t000025	scaffold30010	386795-384065	2	–	–	IIb	AtWRKY42	JcWRKY32
RcWRKY33	30076.t000112	scaffold30076	678304-681837	1	Yes	Yes	IIb	AtWRKY47	JcWRKY30
RcWRKY34	30064.t000028	scaffold30064	205918-203262	–	–	–	IIb	AtWRKY47	JcWRKY33
RcWRKY35	29736.t000019	scaffold29736	182186-187534	–	–	Yes	IIb, misannotated	AtWRKY72	JcWRKY35
RcWRKY36	29822.t000159	scaffold29822	915472-919280	6	–	–	IIb	AtWRKY72	JcWRKY37
RcWRKY37	30147.t000358	scaffold30147	4428301-4432024	–	–	–	IIb	AtWRKY72	JcWRKY36
RcWRKY38	29990.t000001	scaffold29990	4081-6735	–	–	–	IIb	AtWRKY72	JcWRKY34
RcWRKY39	29848.t000020	scaffold29848	96843-94777	46	–	Yes	IId	AtWRKY11	JcWRKY43
RcWRKY40	29598.t000004	scaffold29598	24740-22377	–	–	Yes	IId, misannotated	AtWRKY74	JcWRKY40
RcWRKY41	29644.t000015	scaffold29644	81125-78861	45	Yes	Yes	IId, misassembled	AtWRKY07	JcWRKY41
	29644.t000016	scaffold29644							
RcWRKY42	29883.t000063	scaffold29883	449401-452648	–	–	Yes	IId	AtWRKY21	JcWRKY39
RcWRKY43	30170.t000244	scaffold30170	1224113-1226229	1	–	–	IId	AtWRKY07	JcWRKY42
RcWRKY44	28455.t000009	scaffold28455	137020-134740	–	–	–	IId, misannotated	AtWRKY21	JcWRKY44
RcWRKY45	30174.t000060	scaffold30174	2058419-2060093	9	Yes	Yes	IIe	AtWRKY22	JcWRKY47
RcWRKY46	30190.t000512	scaffold30190	3016527-3018091	–	–	–	IIe	AtWRKY29	JcWRKY45
RcWRKY47	30169.t000358	scaffold30169	2000700-1998672	–	–	–	IIe	AtWRKY27	JcWRKY48
RcWRKY48	30076.t000212	scaffold30076	1385658-1383503	–	–	Yes	IIe	AtWRKY69	JcWRKY46
RcWRKY49	30026.t000025	scaffold30026	178328-180279	2	–	–	IIe	AtWRKY69	JcWRKY50
RcWRKY50	27996.t000002	scaffold27996	6705-10469	–	–	Yes	IIe, misannotated	AtWRKY35	JcWRKY49

(continued)

Table 8.7 (continued)

Gene	Locus ID	Scaffold	Identified position	EST hits	AS[a]	AS[b]	(Sub)group and comments	At_ortholog	Jc_ortholog
RcWRKY51	30190.t000050	scaffold30190	3508665-3505132	2	Yes	Yes	III	AtWRKY41	JcWRKY54
RcWRKY52	30174.t000048	scaffold30174	1952428-1954708	–	–	–	III	AtWRKY30	JcWRKY51
RcWRKY53	30169.t000310	scaffold30169	1952662-1954347	–	–	Yes	III	AtWRKY41	JcWRKY53
RcWRKY54	28690.t000001	scaffold28690	1742258-1744988	–	–	–	III, misannotated	AtWRKY41	JcWRKY52
RcWRKY55	29729.t000063	scaffold29729	13973-11973	–	–	–	III	AtWRKY55	JcWRKY55
RcWRKY56	29729.t000103	scaffold29729	344520-341549	–	–	–	III	AtWRKY55	JcWRKY55
RcWRKY57	29729.t000102	scaffold29729	570667-568714	1	–	Yes	III	AtWRKY70	JcWRKY57
RcWRKY58	29915.t000015	scaffold29915	563637-565065	–	Yes	Yes	III	AtWRKY70	JcWRKY56

[a]Based on the EST data
[b]Based on the RNA sequencing data

subgroup IId members, one subgroup IIa member, one subgroup IIb member, one subgroup IIc member, one subgroup IIe member, and one Group III member) were significantly up-regulated, whereas only two (one subgroup IIe member and one Group III member) were down-regulated. *RcWRKY41*, the most up-regulated gene (more than 250-folds) (Fig. 8.15), was highly expressed in germinating seed, leaf, and male flower (Fig. 8.14), which is consistent with its high representative in Genbank EST database (Table 8.7); its ortholog *AtWRKY11* in *Arabidopsis* was constitutively expressed and acts as negative regulators of basal resistance to *Pseudomonas syringae* (Journot-Catalino et al. 2006). *RcWRKY28*, the second highly up-regulated gene (more than 15-folds) (Table 8.7), was expressed more in male flower and germinating seed than in leaf and endosperm, though its expression level was considerably lower in stage V/VI endosperm as compared to stage II/III (Fig. 8.14); *AtWRKY40*, its ortholog in *Arabidopsis*, was also induced by ABA and acts as a transcriptional repressor in ABA signaling and abiotic stress but a positive regulator in effector-triggered immunity (Shang et al. 2010; Pandey et al. 2010; Schön et al. 2013; Van Aken et al. 2013; Liu et al. 2013). *RcWRKY17*, a Group IIc member preferring to express in male flower, female flower, and germinating seeds, was up-regulated for more than nine folds upon the ABA application (Fig. 8.14); its ortholog *AtWRKY75* in *Arabidopsis* was shown to response to phosphate starvation, water deprivation, ethylene stimulus, and biotic stress, and participate in lateral root development, leaf senescence, and galactolipid biosynthesis (Devaiah et al. 2007; Encinas-Villarejo et al. 2009; Li et al. 2012b; Chen et al. 2013). *RcWRKY45*, a Group IIe member preferring to express in germinating seeds and fruits at 50 days post-anthesis (Li et al. 2012a), was up-regulated for more than sevenfolds by ABA; *AtWRKY22*, its ortholog in *Arabidopsis*, was involved in dark-induced leaf senescence and submergence-mediated immunity (Zhou et al. 2011; Hsu et al. 2013). These results suggested a potential role of RcWRKYs in ABA signaling.

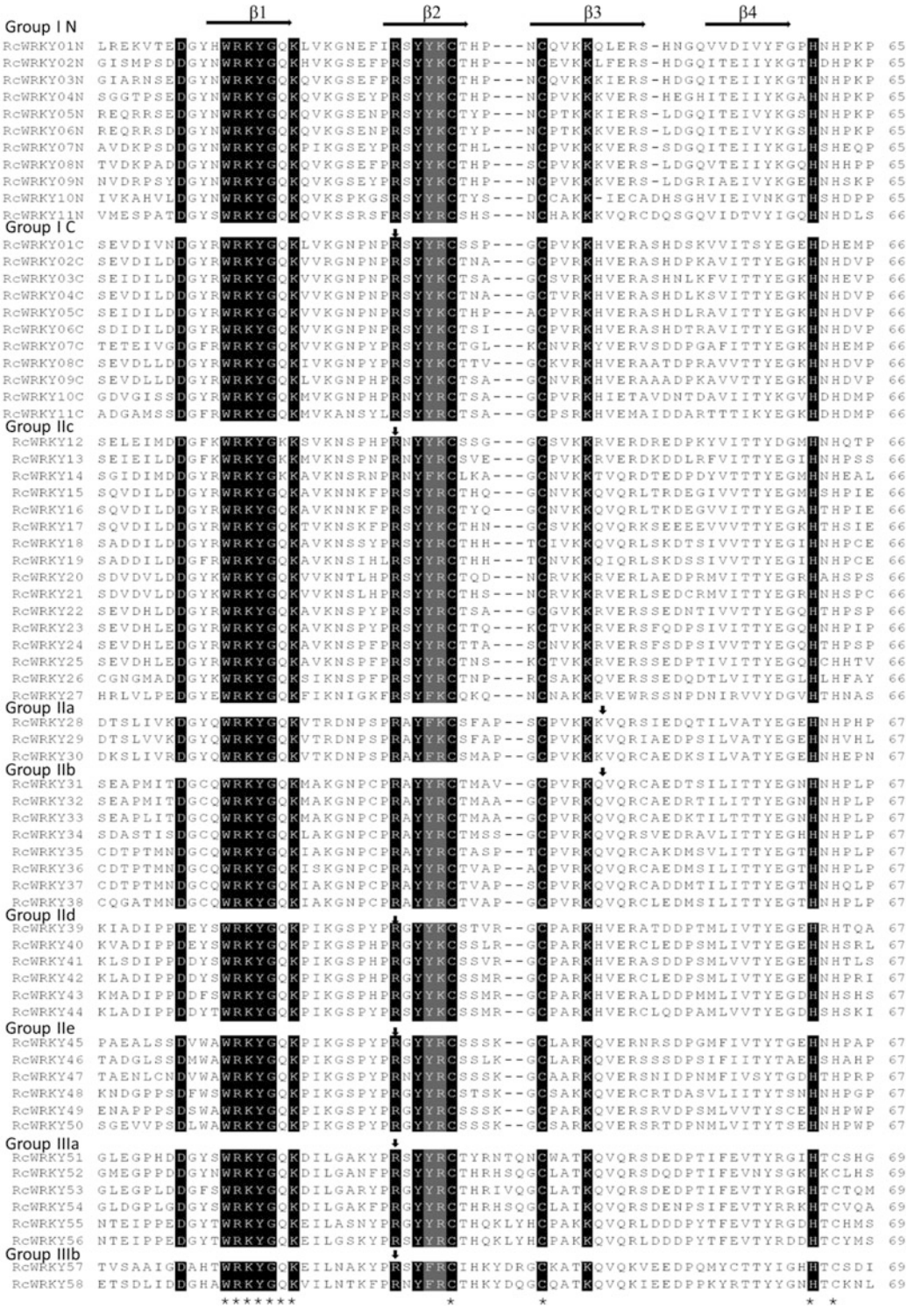

Fig. 8.11 Comparison of WRKY domain sequences of 58 RcWRKY proteins. WRKY••N/C represents the N or C-terminal WRKY domain of Group I members, respectively. Conserved amino acid residues are shown in gray, and the highly conserved WRKYGQ/KK heptapeptide and C_2H_2/C and residues are indicated by "*". The four β-strands are indicated by right arrows. For each (sub)-group, the position of a conserved intron is indicated by a down arrow (after Zou et al. 2016b)

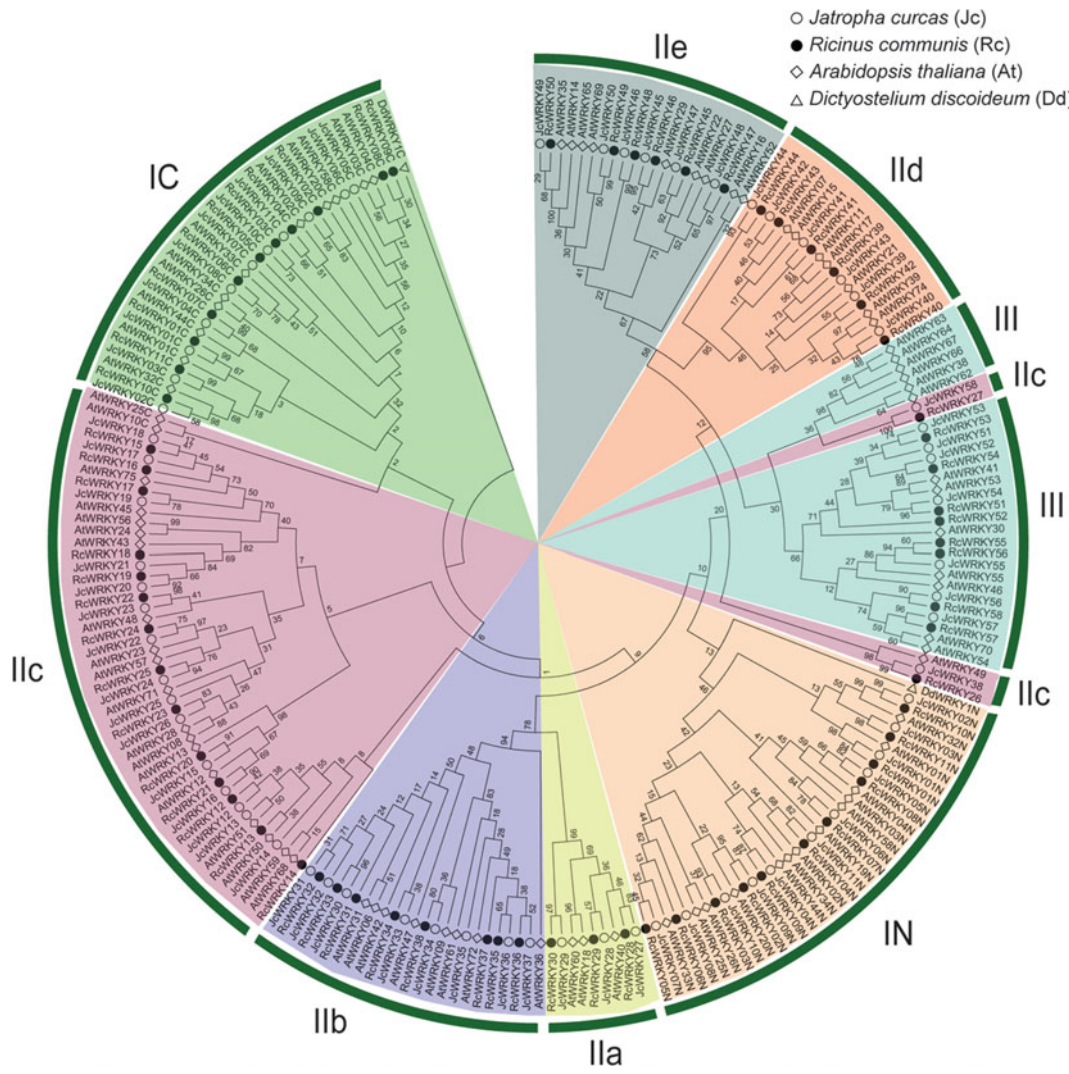

Fig. 8.12 Phylogenetic analysis of castor, physic nut, and *Arabidopsis* WRKY proteins. The WRKY domains (WRKY••N/C representing the N- and C-termini of Group I members, respectively) extracted from deduced amino acid sequences were aligned using MUSCLE. The phylogenetic tree was constructed using the bootstrap neighbor-joining tree (1000 replicates) method implemented in MEGA6, where DdWRKY1C was used as outgroup. The distance scale denotes the number of amino acid substitutions per site. The name of each (sub)group is indicated next to the corresponding group (after Zou et al. 2016b)

8.6 Dof Transcription Factor Gene Family

DNA binding with one finger (Dof) transcription factors, defined by the presence of the Dof domain, comprises a plant-specific family and plays diverse roles in gene regulation of many plant processes such as light-mediated regulation, vascular system development, flowering control, flower abscission, pollen development, endosperm development, seed development and germination, carbon and nitrogen metabolism, and hormone and stress responses (Yanagisawa 2002; Noguero et al. 2013; Gupta et al. 2015; Zou et al. 2019c). The highly conserved Dof

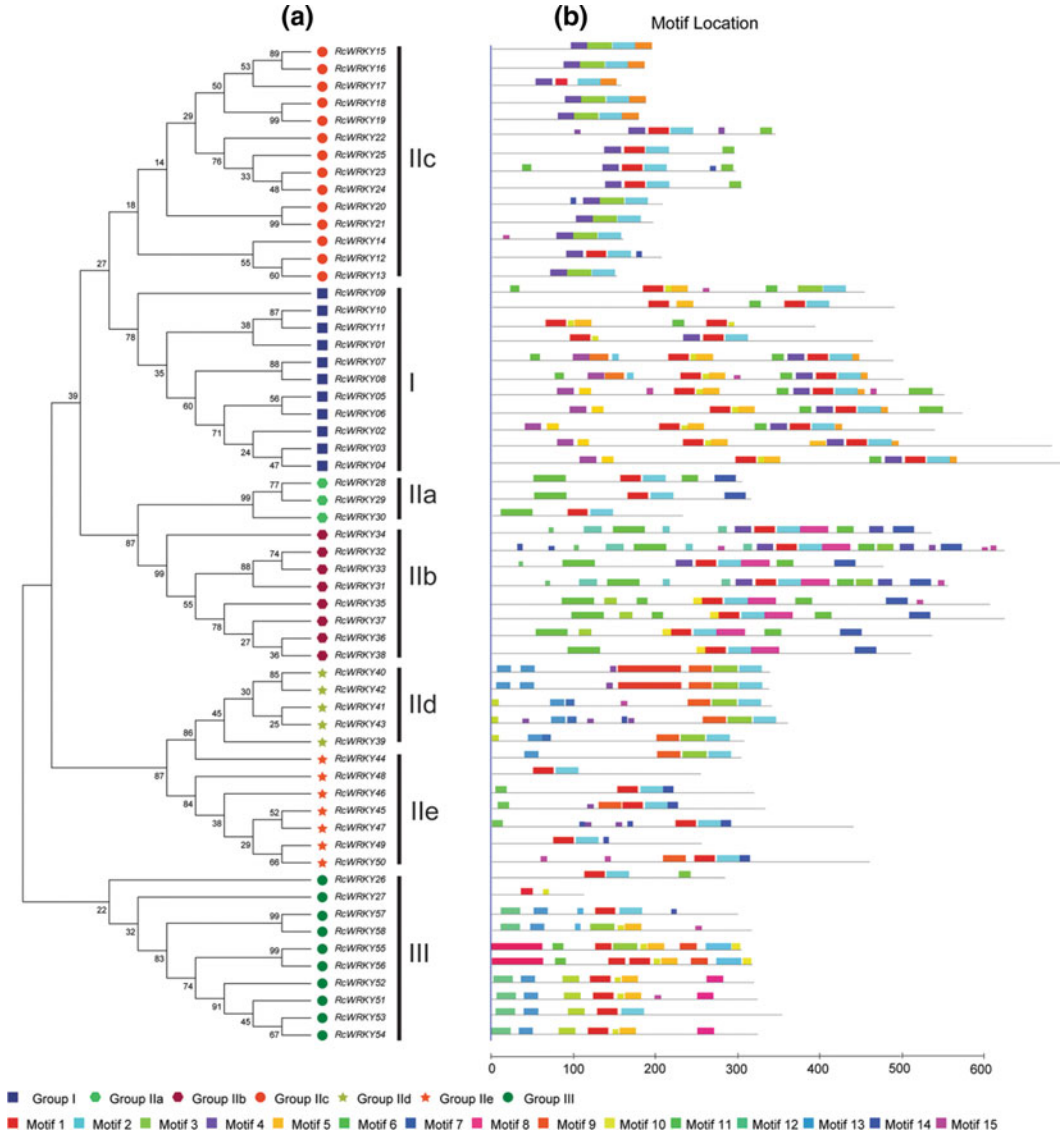

Fig. 8.13 Structural and phylogenetic analysis of RcWRKY proteins. **a** Unrooted phylogenetic tree constructed with full-length RcWRKY proteins and MEGA6. **b** Distribution of conserved motifs among WRKY proteins (motifs are color-coded, as indicated at the bottom of the figure) (after Zou et al. 2016b)

domain, which includes approximately 52 amino acids with a C_2C_2 zinc finger structure, is usually located at the N terminus and is essential for specific binding to the consensus cis-acting element (5′-T/AAAAG-3′) as well as interacting with other proteins such as Ocs element binding factor (OBF) and TEOSINTE BRANCHED1, CYCLOIDEA, and PROLIFERATING CELL NUCLEAR ANTIGEN FACTOR 1 and 2 transcription factor 14 (TCP14) (Yanagisawa and Schmidt 1999; Rueda-Romero et al. 2012; Gupta et al. 2015). Additionally, Dof proteins also contain a bipartite nuclear localization signal (NLS) that partially overlaps with the Dof domain, and a C-terminal transcriptional activation domain (Yanagisawa 2002; Krebs et al. 2010). By contrast, the C-terminal region of Dof proteins is highly variable. Since the first Dof

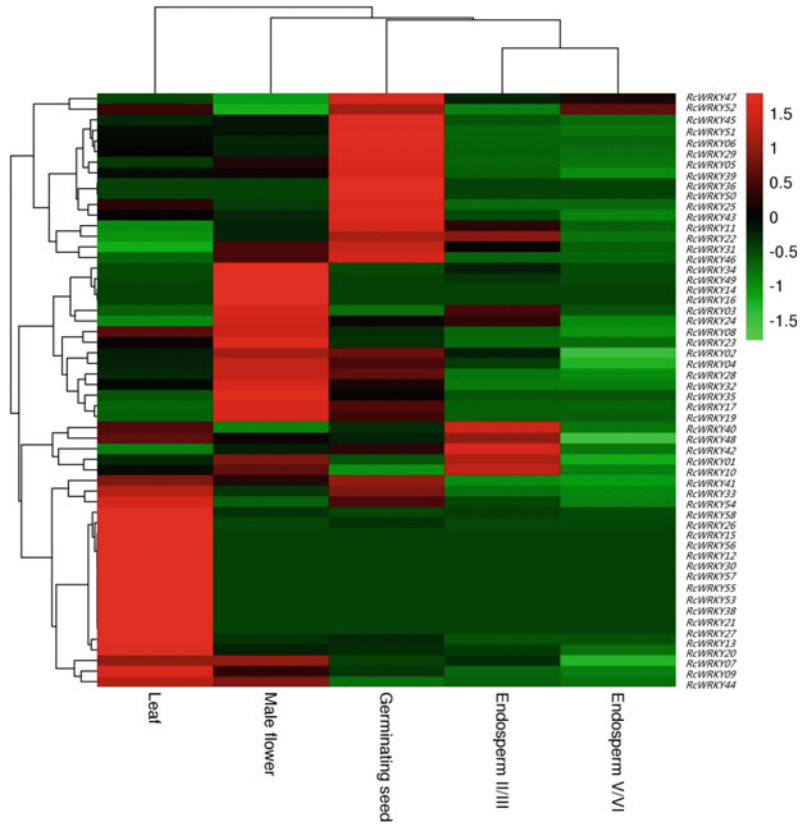

Fig. 8.14 Tissue-specific expression profiles of RcWRKY genes. Color scale represents FPKM normalized $\log_{10}$ transformed counts where green indicates low expression and red indicates high expression (after Zou et al. 2016b)

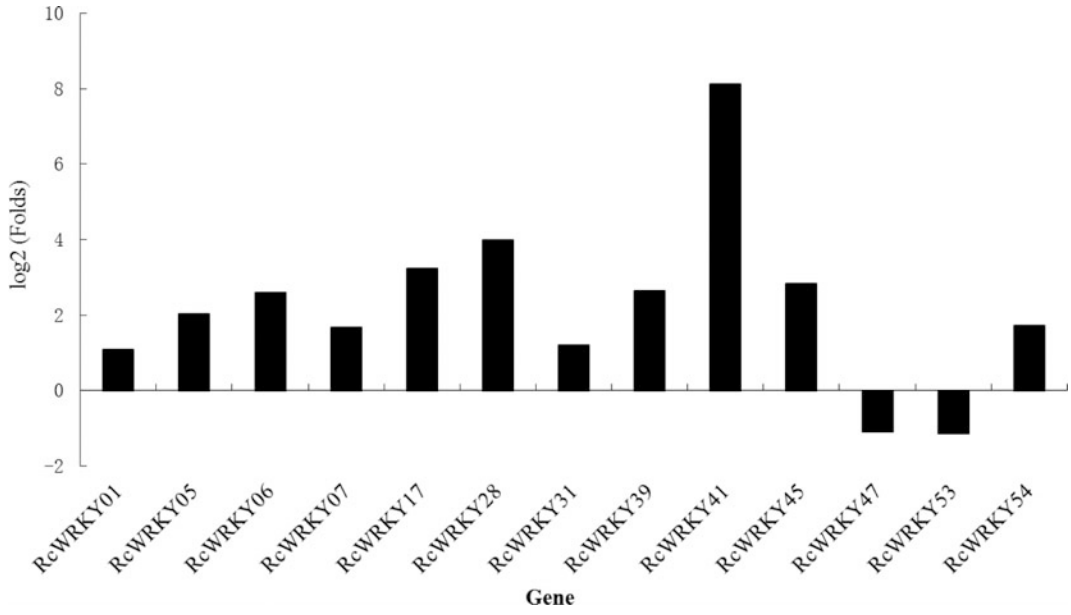

Fig. 8.15 RcWRKY genes associated with ABA response. Developing seeds cultured in vitro were treated with 0 or 10 μM ABA, and samples were collected after the application of 24 h. Differentially expressed genes were identified using Illumina RNA-seq (after Zou et al. 2016b)

Table 8.8 Castor bean Dof family genes (after Zou et al. 2019c)

Gene name	Locus ID	Scaffold location	Nucleotide length (bp, from start to stop codons)		Intron no.	EST no.	Dof location	Comment
			CDS	Gene				
RcDof01	30138.t000239	30138:1637304-1634642	1467	2663	1	5	129-187	–
RcDof02	29813.t000055	29813:341568-343736	1473	2269	1	0	138-196	Misannotated
RcDof03	28838.t000008	28838:42548-39673	1494	2876	1	1	136-194	–
RcDof04	30072.t000004	30072:31747-29003	1521	2745	1	0	110-168	–
RcDof05	29900.t000086	29900:537300-536854	489	897	1	1	36-94	Misannotated
RcDof06	30170.t000353	30170:1886017-1884911	744	744	0	0	25-83	Misannotated
RcDof07	29676.t000058	29676:352074-351337	738	738	0	1	28-86	–
RcDof08	29851.t000017	29851:83027-82104	924	924	0	1	45-103	–
RcDof09	29235.t000001	29235:11382-12419	1038	1038	0	2	34-92	–
RcDof10	29676.t000039	29676:263441-262434	1008	1008	0	0	66-124	–
RcDof11	27395.t000003	27395:18038-17286	753	753	0	1	18-76	–
RcDof12	30138.t000092	30138:654539-655480	942	942	0	3	32-90	–
RcDof13	27985.t000022	27985:152767-153651	912	912	0	0	29-87	Misannotated
RcDof14	29333.t000010	29333:68913-67987	927	927	0	0	10-68	–
RcDof15	30190.t000567	30190:261178-262436	1074	1259	1	1	54-112	–
RcDof16	30190.t000570	30190:276489-275545	945	945	0	5	26-84	–
RcDof17	30147.t000216	30147:3515510-3513099	1005	1656	1	0	52-110	Misannotated
RcDof18	29688.t000002	29688:19134-18280	1047	1934	1	0	82-140	Misannotated
RcDof19	29742.t000016	29742:85032-84102	759	879	1	0	21-79	Misannotated
RcDof20	30098.t000004	30098:40138-41558	897	1421	1	0	40-98	–
RcDof21	30068.t000036	30068:226227-227663	921	1437	1	0	44-102	–
RcDof22	30003.t000001	30003:4571-5770	1017	1200	1	0	77-135	–
RcDof23	30025.t000009	30025:81058-82256	1065	1253	1	0	84-142	Misannotated
RcDof24	28986.t000003	28986:16337-17513	933	1152	1	0	72-130	Misannotated

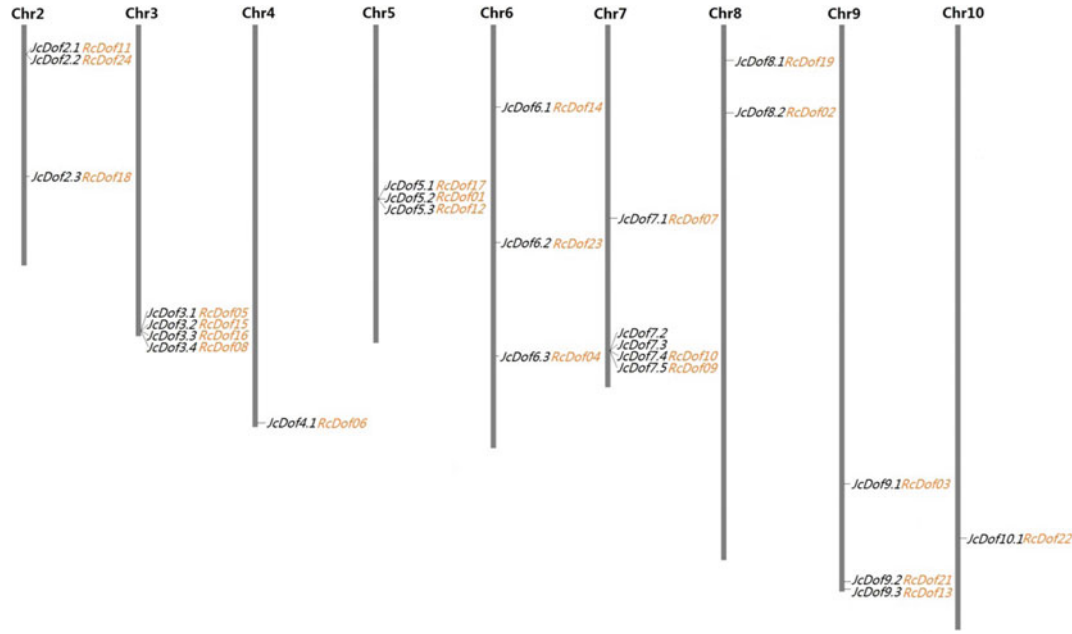

Fig. 8.16 Anchoring RcDof genes to nine physic nut chromosomes. Chromosome serial numbers are indicated at the top of each chromosome, and 23 RcDof genes are shown just next to their collinear genes in physic nut (after Zou et al. 2019c)

gene (*ZmDof1*) was isolated from maize, homologs have been identified in a wide range of plant species. The Dof gene family has been vastly expanded in terrestrial plants relative to most algae, which contain a single Dof gene (Lijavetzky et al. 2003; Moreno-Risueno et al. 2007; Ma et al. 2015).

A survey of the automatic annotation of the castor bean genome identified 21 Dof genes (Jin et al. 2014a, b). However, the manual revision revealed a total of 24 RcDof genes and their expression was all supported by available ESTs and/or RNA-seq reads (Zou et al. 2019c). Furthermore, nine out of the computationally predicted gene models could be optimized (Table 8.8). The 24 RcDof genes are distributed across 21 scaffolds, most of which contain a single Dof gene, whereas scaffold30138, scaffold30190, and scaffold29676 harbor two. Synteny analysis allowed the anchoring of 23 RcDof genes to nine physic nut chromosomes (Fig. 8.16), providing further support of their functionality.

The number of Dof family genes in castor bean is comparable to 25 identified in physic nut and somewhat lower than 30, 36, 41, 45, and 76 found in rice, *Arabidopsis*, poplar, cassava, or Chinese cabbage, respectively (Lijavetzky et al. 2003; Yang et al. 2006; Ma et al. 2015; Zou et al. 2019b). As observed in other plants, RcDof genes contain a single intron or are intronless. All deduced protein sequences were shown to harbor a single Dof domain with the $Cx_2Cx_{21}Cx_2C$ zinc finger pattern and were predicted to contain a nuclear localization signal by using CELLO (http://cello.life.nctu.edu.tw/) (Fig. 8.17).

In rice, *Arabidopsis*, poplar, and Chinese cabbage, Dof proteins have been phylogenetically classified into four subfamilies (i.e. A, B, C, and D), and subfamilies B, C, and D can be further subdivided into several groups (i.e., B1 and B2; C1, C2.1, and C2.2; D1 and D2) (Lijavetzky et al. 2003; Ma et al. 2015). RcDof genes can be divided into nine groups representing four described subfamilies. Groups I, II, IV, V, VII, VIII, and IX correspond to D1, D2, A, C2.2, C1, C2.1, or B1, respectively, whereas Groups III and VI correspond to B2 as described before (Lijavetzky et al. 2003). In fact, the proposed classification is supported by exon–intron structures and conserved motifs. Groups II–V

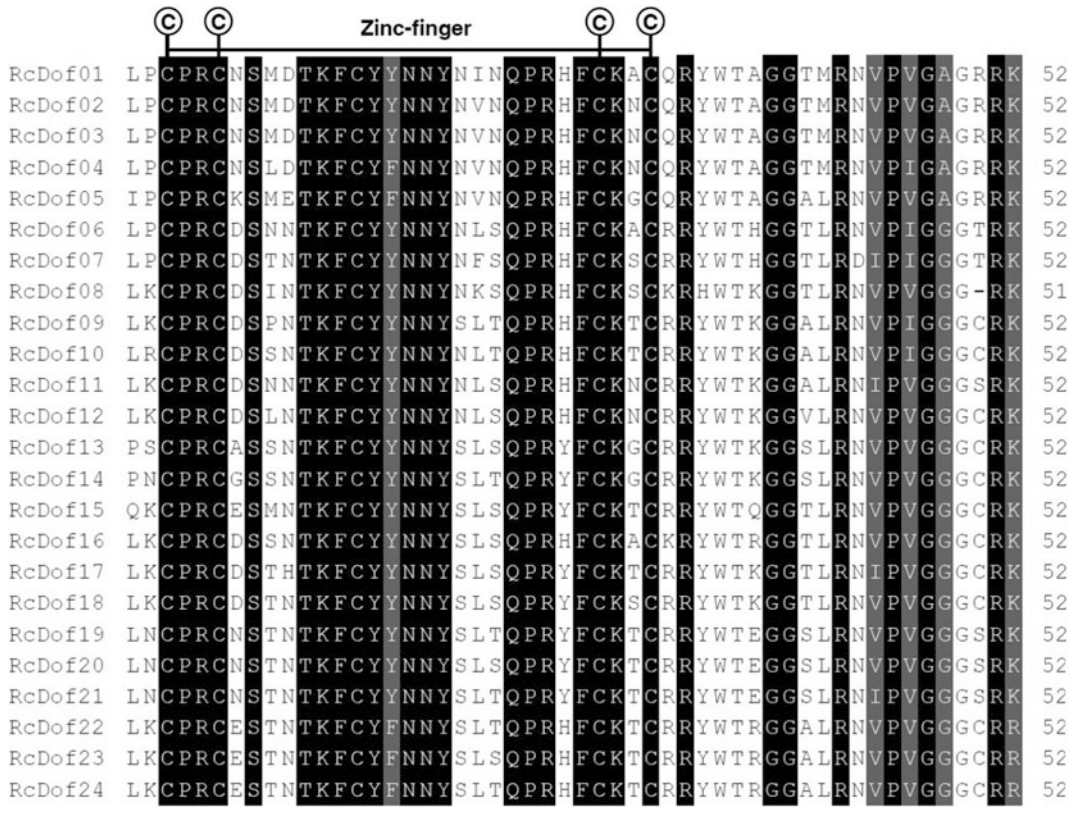

Fig. 8.17 Dof domain sequence alignment of 24 RcDof proteins. The four cysteine residues putatively responsible of the zinc finger structure are indicated. Identical and similar amino acids are highlighted in black or dark gray, respectively (after Zou et al. 2019c)

usually contain no intron, whereas Groups I, VI, VII, VIII, and IX feature one intron with the exception of *RcDof16*. Since the ortholog of *RcDof16* in physic nut (*JcDof3.3*) does not contain an intron either, Group VII can be subdivided into two subgroups. Group IV can also be subdivided into two subgroups (I and II) based on conserved motifs. Compared with subgroup II (including RcDof11 and JcDof5.3), subgroup I (including RcDof11 and JcDof2.1) contains one more motif (Motif 10) (Fig. 8.18).

When using FPKM to represent gene expression levels, transcriptional profiling revealed that the total transcripts of RcDof family genes are most abundant in leaf (Class I); followed by male flower and developing seed (Class II); moderate in endosperm I/III and germinating seed (Class III); and relatively low in endosperm V/VI (Class IV). Classes II, III, and IV accounted for 79.3–79.4, 38.2–56.0, and 13.7% of Class I, respectively. According to their expression patterns, RcDof genes can be grouped into seven main clusters: Cluster I is preferentially expressed in germinating seed, including *RcDof04* and *RcDof12*, which were also shown to be highly expressed in male flower and endosperm I/III, respectively. Cluster II is mostly expressed in developing seed, including *RcDof21*, *RcDof08*, *RcDof11*, *RcDof16,* and *RcDof20*. Cluster III shows expression in endosperm I/III, including *RcDof07*, *RcDof19*, *RcDof06,* and *RcDof18*. *RcDof19*, *RcDof06,* and *RcDof18* also showed high expression in developing seed. Cluster IV contains *RcDof13* and *RcDof05* that are preferentially expressed in male flower. Cluster V comprises *RcDof03*, *RcDof01*, and *RcDof09* that are preferentially expressed in leaf and male flower. *RcDof10*, *RcDof22,* and *RcDof23*

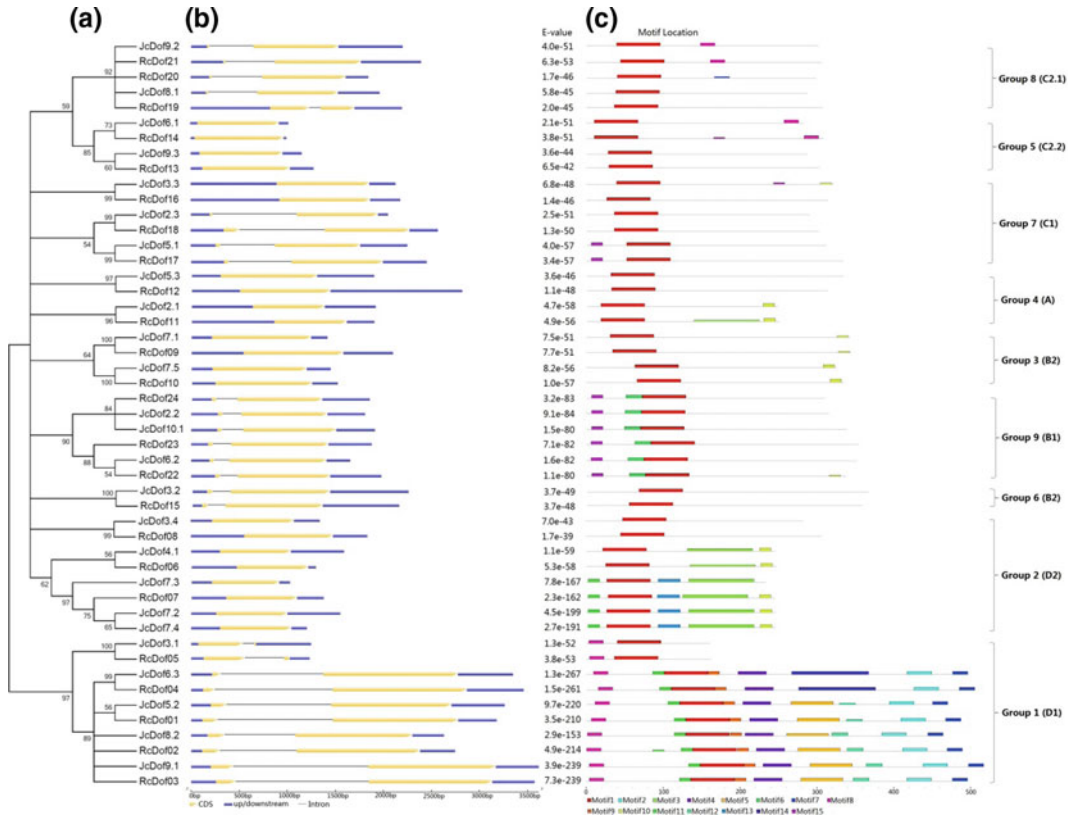

Fig. 8.18 Structural and phylogenetic analysis of physic nut and castor bean Dof genes. **a** Unrooted phylogenetic tree was constructed using full-length Dof proteins with MEGA6. **b** Graphic representation of exon–intron structures was displayed using GSDS (http://gsds.cbi.pku.edu.cn/). **c** Distribution of conserved motifs among Dof proteins (motifs are color-coded, as indicated at the bottom of the figure) (after Zou et al. 2019c)

compose Cluster VI, showing expression in leaf and developing seed. Cluster VII is expressed in leaf and includes *RcDof24*, *RcDof02*, *RcDof15,* and *RcDof17* (Fig. 8.19).

Studies have shown that Dof genes are widely participated in ABA and GA (gibberellic acid) signal pathways (Gualberti et al. 2002; Moreno-Risueno et al. 2007; Gabriele et al. 2010), implying that the functions of Dof genes might be mediated by ABA and GA signals. To dissect the possible involvement of RcDof genes in the regulation of hormone responses, the expression levels of RcDof genes were investigated in response to ABA and GA treatments, respectively. Semi-quantitative RT-PCR analysis of three-leaf stage seedlings revealed that 18 RcDof genes were highly regulated in leaves after application of 100 μM ABA or GA for 40 h: 13 genes including *RcDof08* were up-regulated, and two genes were down-regulated by ABA; seven genes were up-regulated, and five genes were down-regulated by GA; seven GA-inducible genes were also up-regulated by ABA; however, *RcDof15* was up-regulated by ABA but down-regulated by GA (Jin et al. 2014a, b). In developing seeds cultured in vivo, three genes (*RcDof01*, *RcDof08,* and *RcDof18*) were significantly up-regulated after application of 10 μM ABA for 24 h (Chandrasekaran et al. 2014) (Fig. 8.20). ABA-inducible expression of *RcDof08* in both leaf and seed implies its key role in ABA signaling.

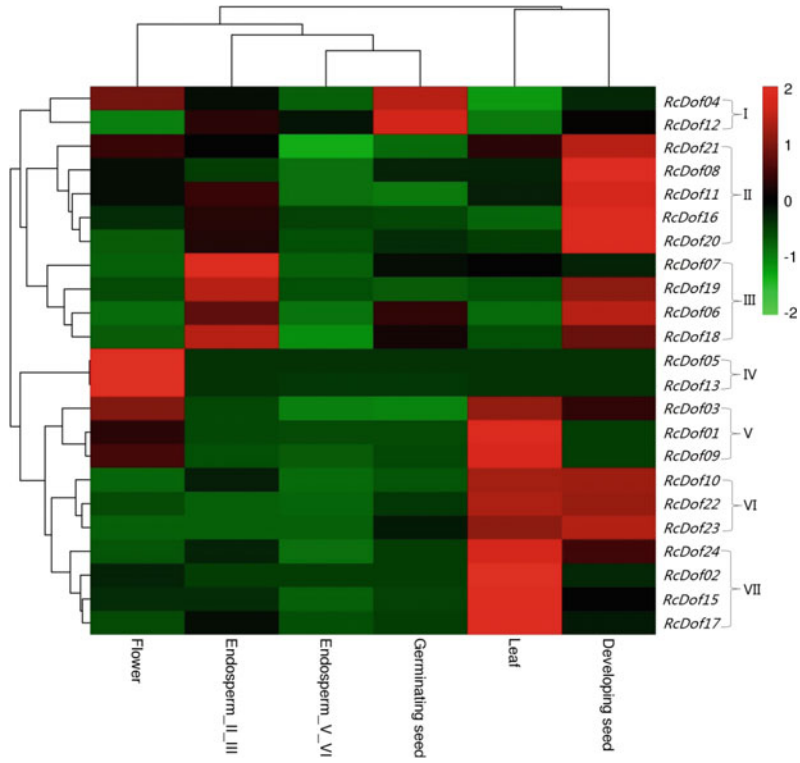

Fig. 8.19 Tissue-specific expression profiles of RcDof genes. Color scale represents FPKM normalized $\log_{10}$ transformed counts where green indicates low expression and red indicates high expression (after Zou et al. 2019c)

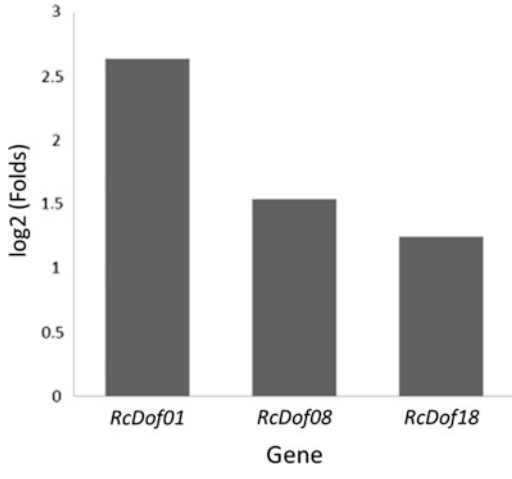

Fig. 8.20 RcDof genes associated with ABA response in developing seeds cultured in vivo. Developing seeds cultured in vitro were treated with 0 or 10 μM ABA, and samples were collected after the application of 24 h. Differentially expressed genes were identified using Illumina RNA-seq (after Zou et al. 2019c)

8.7 Concluding Remarks and Future Prospects

WGD results in polyploidy, which is an important evolutionary mechanism for generation of new genes. The growing number of sequenced plant genomes has showed that WGD is widespread. More than 30 WGD events have been described in different plant lineages, including the ancient γ event in core eudicots as well as the so-called τ event in monocots (Jiao et al. 2014; Vanneste et al. 2014; Ren et al. 2018). Moreover, two representative model plants, *Arabidopsis* and rice, experienced two recent whole-genome doubling events, followed by massive gene losses and chromosomal rearrangements (Bowers et al. 2003). By contrast, no additional WGD and limited gene loss occurred in castor bean after the γ event. From this perspective, mining gene

families in castor bean will improve our understanding of lineage-specific gene expansion and evolution in higher plants, especially in core eudicots. In order to advance our knowledge of castor bean gene families, additional and improved genomic resources are needed. The castor bean genome sequence assembly is currently fragmented in 25,763 scaffolds with an N50 (a measure of a set of scaffold lengths, which was defined as the shortest sequence length at 50% of the genome) of 21,409 bp. Efforts to construct a high-density genetic map are ongoing. Additional sequence data, including those produced with new sequencing platforms that generate contiguous sequence reads up to tens of kbp-long, will largely improve the castor bean genome sequence and will facilitate the anchoring of sequence assemblies to the physical chromosomes.

Genome annotation could also be improved with additional transcriptomic analyses of multiple tissues, developmental stages, and environmental conditions.

Finally, functional analysis of genes identified in the genome would be dramatically accelerated with the construction of mutant collections and the development of efficient genetic transformation platforms.

Acknowledgements This work was supported by the National Natural Science Foundation of China (31700580), the Natural Science Foundation of Hainan Province (312026), and the Central Public-interest Scientific Institution Basal Research Fund for Chinese Academy of Tropical Agricultural Sciences (1630022019017, 1630022011014 and 1630022017022).

References

Abascal F, Irisarri I, Zardoya R (2014) Diversity and evolution of membrane intrinsic proteins. Biochim Biophys Acta 1840:1468–1481

Ahmed SU, Rojo E, Kovaleva V et al (2000) The plant vacuolar sorting receptor AtELP is involved in transport of NH(2)-terminal propeptide-containing vacuolar proteins in *Arabidopsis thaliana*. J Cell Biol 149:1335–1344

Anderberg HI, Danielson JÅ, Johanson U (2011) Algal MIPs, high diversity and conserved motifs. BMC Evol Biol 11:110

Anderberg HI, Kjellbom P, Johanson U (2012) Annotation of *Selaginella moellendorffii* major intrinsic proteins and the evolution of the protein family in terrestrial plants. Front Plant Sci 3:33

Beers EP, Jones AM, Dickerman AW (2004) The S8 serine, C1A cysteine and A1 aspartic protease families in *Arabidopsis*. Phytochemistry 65:43–58

Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422:433–438

Brown AP, Kroon JT, Swarbreck D, Febrer M, Larson TR, Graham IA, Caccamo M, Slabas AR (2012) Tissue-specific whole transcriptome sequencing in castor, directed at understanding triacylglycerol lipid biosynthetic pathways. PLoS ONE 7:e30100

Chan AP, Crabtree J, Zhao Q et al (2010) Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotechnol 28:951–956

Chandrasekaran U, Liu A (2014) Endogenous abscisic acid signaling towards storage reserve filling in developing seed tissues of castor bean (*Ricinus communis* L.). Plant Growth Regul 72:203–207

Chandrasekaran U, Xu W, Liu A (2014) Transcriptome profiling identifies ABA mediated regulatory changes towards storage filling in developing seeds of castor bean (*Ricinus communis* L.). Cell Biosci 4:33

Chen H, Lai Z, Shi J, Xiao Y, Chen Z, Xu X (2010) Roles of arabidopsis WRKY18, WRKY40 and WRKY60 transcription factors in plant responses to abscisic acid and abiotic stress. BMC Plant Biol 10:281

Chen L, Song Y, Li S, Zhang L, Zou C, Yu D (2012) The role of WRKY transcription factors in plant abiotic stresses. Biochim Biophys Acta 1819:120–128

Chen X, Liu J, Lin G, Wang A, Wang Z, Lu G (2013) Overexpression of *AtWRKY28* and *AtWRKY75* in *Arabidopsis* enhances resistance to oxalic acid and *Sclerotinia sclerotiorum*. Plant Cell Rep 32:1589–1599

Chileh T, Esteban-García B, Alonso DL, García-Maroto F (2010) Characterization of the 11S globulin gene family in the castor plant *Ricinus communis* L. J Agric Food Chem 58:272–281

Croce R, van Amerongen H (2014) Natural strategies for photosynthetic light harvesting. Nat Chem Biol 10:492–501

Danielson JA, Johanson U (2008) Unexpected complexity of the aquaporin gene family in the moss *Physcomitrella patens*. BMC Plant Biol 8:45

Devaiah BN, Karthikeyan AS, Raghothama KG (2007) WRKY75 transcription factor is a modulator of phosphate acquisition and root development in *Arabidopsis*. Plant Physiol 143:1789–1801

Doering-Saad C, Newbury HJ, Couldridge CE et al (2006) A phloem-enriched cDNA library from *Ricinus*: insights into phloem function. J Exp Bot 57:3183–3193

Drenth J, Jansonius JN, Koekoek R, Swen HM, Wolthers BG (1968) Structure of papain. Nature 218:929–932

Eisenbarth DA, Weig AR (2005) Dynamics of aquaporins and water relations during hypocotyl elongation in *Ricinus communis* L. seedlings. J Exp Bot 56:1831–1842

Encinas-Villarejo S, Maldonado AM, Amil-Ruiz F et al (2009) Evidence for a positive regulatory role of strawberry (*Fragaria* × *ananassa*) FaWRKY1 and *Arabidopsis* AtWRKY75 proteins in resistance. J Exp Bot 60:3043–3065

Engelken J, Brinkmann H, Adamska I (2010) Taxonomic distribution and origins of the extended LHC (light-harvesting complex) antenna protein superfamily. BMC Evol Biol 10:233

Eulgem T, Rushton PJ, Robatzek S, Somssich IE (2000) The WRKY superfamily of plant transcription factors. Trends Plant Sci 5:199–206

Froger A, Tallur B, Thomas D, Delamarche C (1998) Prediction of functional residues in water channels and related proteins. Protein Sci 7:1458–1468

Gabriele S, Rizza A, Martone J, Circelli P, Costantino P, Vittorioso P (2010) The Dof protein DAG1 mediates PIL5 activity on seed germination by negatively regulating GA biosynthetic gene AtGA3ox1. Plant J 61:312–323

Gaut BS, Morton BR, McCaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene rbcL. Proc Natl Acad Sci USA 93:10274–10279

Gomes D, Agasse A, Thiébaud P, Delrot S, Gerós H, Chaumont F (2009) Aquaporins are multifunctional water and solute transporters highly divergent in living organisms. Biochim Biophys Acta 1788:1213–1228

Gualberti G, Papi M, Bellucci L, Ricci I, Bouchez D, Camilleri C, Costantino P, Vittorioso P (2002) Mutations in the Dof zinc finger genes DAG2 and DAG1 influence with opposite effects the germination of Arabidopsis seeds. Plant Cell 14:1253–1263

Guo C, Guo R, Xu X, Gao M, Li X, Song J, Zheng Y, Wang X (2014) Evolution and expression analysis of the grape (*Vitis vinifera* L.) WRKY gene family. J Exp Bot 65:1513–1528

Gupta S, Malviya N, Kushwaha H, Nasim J, Bisht NC, Singh VK, Yadav D (2015) Insights into structural and functional diversity of Dof (DNA binding with one finger) transcription factor. Planta 241:549–562

He H, Dong Q, Shao Y, Jiang H, Zhu S, Cheng B, Xiang Y (2012) Genome-wide survey and characterization of the WRKY gene family in *Populus trichocarpa*. Plant Cell Rep 31:1199–1217

Hierl G, Vothknecht U, Gietl C (2012) Programmed cell death in *Ricinus* and *Arabidopsis*: the function of KDEL cysteine peptidases in development. Physiol Plant 145:103–113

Hove RM, Bhave M (2011) Plant aquaporins with non-aqua functions: deciphering the signature sequences. Plant Mol Biol 75:413–430

Hsu FC, Chou MY, Chou SJ, Li YR, Peng HP, Shih MC (2013) Submergence confers immunity mediated by the WRKY22 transcription factor in *Arabidopsis*. Plant Cell 25:2699–2713

Jansson S (1999) A guide to the *Lhc* genes and their relatives in *Arabidopsis*. Trends Plant Sci 4:236–240

Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR et al (2012) A genome triplication associated with early diversification of the core eudicots. Genome Biol 13:R3

Jiao Y, Li J, Tang H, Paterson AH (2014) Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. Plant Cell 26:2792–2802

Jin Z, Chandrasekaran U, Liu A (2014a) Genome-wide analysis of the Dof transcription factors in castor bean (*Ricinus communis* L.). Genes Genom 36:527–537

Jin Z, Xu W, Liu A (2014b) Genomic surveys and expression analysis of bZIP gene family in castor bean (*Ricinus communis* L.). Planta 239:299–312

Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucl Acids Res 45(D1):D1040–D1045

Journot-Catalino N, Somssich IE, Roby D, Kroj T (2006) The transcription factors WRKY11 and WRKY17 act as negative regulators of basal resistance in *Arabidopsis thaliana*. Plant Cell 18:3289–3302

Kirscht A, Kaptan SS, Bienert GP, Chaumont F, Nissen P, de Groot BL, Kjellbom P, Gourdon P, Johanson U (2016) Crystal structure of an ammonia-permeable aquaporin. PLoS Biol 14(3):e1002411

Klimmek F, Sjödin A, Noutsos C, Leister D, Jansson S (2006) Abundantly and rarely expressed Lhc protein genes exhibit distinct regulation patterns in plants. Plant Physiol 140:793–804

Koziol AG, Borza T, Ishida K, Keeling P, Lee RW, Durnford DG (2007) Tracing the evolution of the light-harvesting antennae in chlorophyll a/b-containing organisms. Plant Physiol 143:1802–1816

Krebs J, Mueller-Roeber B, Ruzicic S (2010) A novel bipartite nuclear localization signal with an atypically long linker in DOF transcription factors. J Plant Physiol 67:583–586

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359

Li HL, Zhang LB, Guo D, Li CZ, Peng SQ (2012a) Identification and expression profiles of the WRKY transcription factor family in *Ricinus communis*. Gene 503:248–253

Li Z, Peng J, Wen X, Guo H (2012b) Gene network analysis and functional studies of senescence-associated genes reveal novel regulators of *Arabidopsis* leaf senescence. J Integr Plant Biol 54:526–539

Lijavetzky D, Carbonero P, Vicente-Carbajosa J (2003) Genome-wide comparative phylogenetic analysis of the rice and *Arabidopsis* Dof gene families. BMC Evol Biol 3:17

Liu R, Xu YH, Jiang SC, Lu K, Lu YF, Feng XJ, Wu Z, Liang S, Yu YT, Wang XF, Zhang DP (2013)

Light-harvesting chlorophyll a/b-binding proteins, positively involved in abscisic acid signalling, require a transcription repressor, WRKY40, to balance their function. J Exp Bot 64:5443–5456

Lu CF, Wallis JG, Browse J (2007) An analysis of expressed sequence tags of developing castor endosperm using a full-length cDNA library. BMC Plant Biol 7:42

Lu H, Chandrasekar B, Oeljeklaus J, Misas-Villamil JC, Wang Z, Shindo T, Bogyo M, Kaiser M, van der Hoorn RA (2015) Subfamily-specific fluorescent probes for cysteine proteases display dynamic protease activities during seed germination. Plant Physiol 168:1462–1475

Ma J, Li MY, Wang F, Tang J, Xiong AS (2015) Genome-wide analysis of Dof family transcription factors and their responses to abiotic stresses in Chinese cabbage. BMC Genom 16:33

Martínez M, Diaz I (2008) The origin and evolution of plant cystatins and their target cysteine proteinases indicate a complex functional relationship. BMC Evol Biol 8:198

Maurel C, Verdoucq L, Luu DT, Santoni V (2008) Plant aquaporins: membrane channels with multiple integrated functions. Annu Rev Plant Biol 59:595–624

Misas-Villamil JC, van der Hoorn RA, Doehlemann G (2016) Papain-like cysteine proteases as hubs in plant immunity. New Phytol 212:902–907

Moreno-Risueno MA, Martínez M, Vicente-Carbajosa J, Carbonero P (2007) The family of DOF transcription factors: from green unicellular algae to vascular plants. Mol Genet Genom 277:379–390

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Method 5:621–628

Noguero M, Atif RM, Ochatt S, Thompson RD (2013) The role of the DNA-binding one zinc finger (DOF) transcription factor family in plants. Plant Sci 209:32–45

Ogunniyi DS (2006) Castor oil: a vital industrial raw material. Bioresour Technol 97:1086–1091

Pandey SP, Roccaro M, Schön M, Logemann E, Somssich IE (2010) Transcriptional reprogramming regulated by WRKY18 and WRKY40 facilitates powdery mildew infection of *Arabidopsis*. Plant J 64:912–923

Preston GM, Carroll TP, Guggino WB, Agre P (1992) Appearance of water channels in *Xenopus* oocytes expressing red cell CHIP28 protein. Science 256:385–387

Qiu L, Yang C, Tian B, Yang JB, Liu A (2010) Exploiting EST databases for the development and characterization of EST-SSR markers in castor bean (*Ricinus communis* L.). BMC Plant Biol 10:278

Rawlings ND, Barrett AJ, Finn R (2016) Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. Nucl Acids Res 44:D343–D350

Ren R, Wang H, Guo C, Zhang N, Zeng L, Chen Y, Ma H, Qi J (2018) Wide-spread whole genome duplications contribute to genome complexity and species diversity in angiosperms. Mol Plant 11:414–428

Richau KH, Kaschani F, Verdoes M, Pansuriya TC, Niessen S, Stüber K, Colby T, Overkleeft HS, Bogyo M, Van der Hoorn RA (2012) Subclassification and biochemical analysis of plant papain-like cysteine proteases displays subfamily-specific characteristics. Plant Physiol 158:1583–1599

Rinerson C, Rabara RC, Tripathi P, Shen QJ, Rushton PJ (2015) The evolution of WRKY transcription factors. BMC Plant Biol 15:66

Rivarola M, Foster JT, Chan AP, Williams AL, Rice DW, Liu X, Melake-Berhan A, Huot Creasy H, Puiu D, Rosovitz MJ, Khouri HM, Beckstrom-Sternberg SM, Allan GJ, Keim P, Ravel J, Rabinowicz PD (2011) Castor bean organelle genome sequencing and worldwide genetic diversity analysis. PLoS ONE 6:e21743

Rochaix JD (2014) Regulation and dynamics of the light-harvesting system. Annu Rev Plant Biol 65:287–309

Rueda-Romero P, Barrero-Sicilia C, Gómez-Cadenas A, Carbonero P, Oñate-Sánchez L (2012) *Arabidopsis thaliana* DOF6 negatively affects germination in non-after-ripened seeds and interacts with TCP14. J Exp Bot 63:1937–1949

Rushton PJ, Somssich IE, Ringler P, Shen QJ (2010) WRKY transcription factors. Trends Plant Sci 15:247–258

Schluttenhofer C, Yuan L (2015) Regulation of specialized metabolism by WRKY transcription factors. Plant Physiol 167:295–306

Schmid M, Simpson D, Kalousek F, Gietl C (1998) A cysteine endopeptidase with a C-terminal KDEL motif isolated from castor bean endosperm is a marker enzyme for the ricinosome, a putative lytic compartment. Planta 206:466–475

Schmid M, Simpson D, Gietl C (1999) Programmed cell death in castor bean endosperm is associated with the accumulation and release of a cysteine endopeptidase from ricinosomes. Proc Natl Acad Sci USA 96:14159–14164

Schmid M, Simpson DJ, Sarioglu H, Lottspeich F, Gietl C (2001) The ricinosomes of senescing plant tissue bud from the endoplasmic reticulum. Proc Natl Acad Sci USA 98:5353–5358

Schön M, Töller A, Diezel C, Roth C, Westphal L, Wiermer M, Somssich IE (2013) Analyses of wrky18 wrky40 plants reveal critical roles of SA/EDS1 signaling and indole-glucosinolate biosynthesis for *Golovinomyces orontii* resistance and a loss-of resistance towards *Pseudomonas syringae* pv. tomato AvrRPS4. Mol Plant Microbe Interact 26:758–767

Shang Y, Yan L, Liu ZQ, Cao Z, Mei C et al (2010) The Mg-chelatase H subunit of *Arabidopsis* antagonizes a group of WRKY transcription repressors to relieve ABA-responsive genes of inhibition. Plant Cell 22:1909–1935

Sood A, Jaiswal V, Chanumolu SK, Malhotra N, Pal T, Chauhan RS (2014) Mining whole genomes and

transcriptomes of Jatropha (*Jatropha curcas*) and Castor bean (*Ricinus communis*) for NBS-LRR genes and defense response associated transcription factors. Mol Biol Rep 41:7683–7695

Tan M, Xue J, Wang L, Huang J, Fu C, Yan X (2016) Transcriptomic analysis for different sex types of *Ricinus communis* L. during development from apical buds to inflorescences by digital gene expression profiling. Front Plant Sci 6:1208

Than ME, Helm M, Simpson DJ, Lottspeich F, Huber R, Gietl C (2004) The 2.0 A crystal structure and substrate specificity of the KDEL-tailed cysteine endopeptidase functioning in programmed cell death of *Ricinus communis* endosperm. J Mol Biol 336:1103–1116

Törnroth-Horsefield S, Wang Y, Hedfalk K, Johanson U, Karlsson M, Tajkhorshid E, Neutze R, Kjellbom P (2006) Structural mechanism of plant aquaporin gating. Nature 439:688–694

Troncoso-Ponce MA, Kilaru A, Cao X, Durrett TP, Fan J, Jensen JK, Thrower NA, Pauly M, Wilkerson C, Ohlrogge JB (2011) Comparative deep transcriptional profiling of four developing oilseeds. Plant J 68:1014–1027

Turk V, Turk B, Turk D (2001) Lysosomal cysteine proteases: facts and opportunities. EMBO J 20:4629–4633

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596–1604

Van Aken O, Zhang B, Law S, Narsai R, Whelan J (2013) AtWRKY40 and AtWRKY63 modulate the expression of stress-responsive nuclear genes encoding mitochondrial and chloroplast proteins. Plant Physiol 162:254–271

van de Loo FJ, Broun P, Turner S, Somerville C (1995) An oleate 12-hydroxylase from *Ricinus communis* L. is a fatty acyl desaturase homolog. Proc Natl Acad Sci USA 92:6743–6747

van der Hoorn RA (2008) Plant proteases: from phenotypes to molecular mechanisms. Annu Rev Plant Biol 59:191–223

Vanneste K, Baele G, Maere S, Van de Peer Y (2014) Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. Genome Res 24:1334–1347

Zou Z, Gong J, An F, Xie G, Wang J et al (2015a) Genome-wide identification of rubber tree (*Hevea brasiliensis* Muell. Arg.) aquaporin genes and their response to ethephon stimulation in the laticifer, a rubber-producing tissue. BMC Genomics 16:1001

Xiong W, Xu X, Zhang L, Wu P, Chen Y, Li M, Jiang H, Wu G (2013) Genome-wide analysis of the WRKY gene family in physic nut (*Jatropha curcas* L.). Gene 524:124–132

Xu W, Li F, Ling L, Liu A (2013) Genome-wide survey and expression profiles of the AP2/ERF family in castor bean (*Ricinus communis* L.). BMC Genomics 14:785

Xu W, Dai M, Li F, Liu A (2014) Genomic imprinting, methylation and parent-of-origin effects in reciprocal hybrid endosperm of castor bean. Nucl Acids Res 42:6987–6998

Xu W, Chen Z, Ahmed N, Han B, Cui Q, Liu A (2016) Genome-wide identification, evolutionary analysis, and stress responses of the GRAS gene family in castor beans. Int J Mol Sci 17(pii):E1004

Yanagisawa S (2002) The Dof family of plant transcription factors. Trends Plant Sci 7:555–560

Yanagisawa S, Schmidt RJ (1999) Diversity and similarity among recognition sequences of Dof transcription factors. Plant J 17:209–214

Yang X, Tuskan GA, Cheng MZ (2006) Divergence of the Dof gene families in poplar, Arabidopsis, and rice suggests multiple modes of gene evolution after duplication. Plant Physiol 142:820–830

Zeng C, Wang W, Zheng Y, Chen X, Bo W, Song S, Zhang W, Peng M (2010) Conservation and divergence of microRNAs and their functions in Euphorbiaceous plants. Nucl Acids Res 38:981–995

Zhang SD, Ling LZ (2014) Genome-wide identification and evolutionary analysis of the SBP-box gene family in castor bean. PLoS ONE 9:e86688

Zhang Y, Wang L (2005) The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. BMC Evol Biol 5:1

Zhou X, Jiang Y, Yu D (2011) WRKY22 transcription factor mediates dark-induced leaf senescence in *Arabidopsis*. Mol Cells 31:303–313

Zou Z, Huang QX, An F (2013) Genome-wide identification, classification and expression analysis of Lhc supergene family in castor bean (*Ricinus communis* L.). Agri Biotechnol 2:44–48, 51

Zou Z, Gong J, Huang Q, Mo Y, Yang L, Xie G (2015b) Gene structures, evolution, classification and expression profiles of the aquaporin gene family in castor bean (*Ricinus communis* L.). PLoS One 10: e0141022

Zou Z, Yang L, Gong J, Mo Y, Wang J, Cao J, An F, Xie G (2016a) Genome-wide identification of *Jatropha curcas* aquaporin genes and the comparative analysis provides insights into the gene family expansion and evolution in *Hevea brasiliensis*. Front Plant Sci 7:395

Zou Z, Yang L, Wang D, Huang Q, Mo Y, Xie G (2016b) Gene structures, evolution and transcriptional profiling of the WRKY gene family in castor bean (*Ricinus communis* L.). PLoS One 11:e0148243

Zou Z, Liu JT, Xie GS, Yang LF (2017a) Survey of the rubber tree genome reveals a high number of cysteine protease-encoding genes homologous to *Arabidopsis* SAG12. PLoS ONE 12:e0171725

Zou Z, Xie GS, Yang LF (2017b) Papain-like cysteine protease encoding genes in rubber (*Hevea brasiliensis*): comparative genomics, phylogenetic and transcriptional profiling analysis. Planta 246:999–1018

Zou Z, Huang QX, Xie GS, Yang LF (2018) Genome-wide comparative analysis of papain-like cysteine protease family genes in castor bean and physic nut. Sci Rep 8:331

Zou Z, Yang JH, Zhang XC (2019a) Insights into genes encoding respiratory burst oxidase homologs (RBOHs) in rubber tree (Hevea brasiliensis Muell. Arg.). Ind Crop Prod 128:126–139

Zou Z, Zhang XC (2019c) Genome-wide identification and comparative evolutionary analysis of the Dof transcription factor family in physic nut and castor bean. PeerJ 7:e6354

Zou Z, Zhu JL, Zhang XC (2019b) Genome-wide identification and characterization of the Dof gene family in cassava (Manihot esculenta). Gene 687:298–307

Zwiazek JJ, Xu H, Tan X, Navarro-Ródenas A, Morte A (2017) Significance of oxygen transport through aquaporins. Sci Rep 7:40411

Sexual Reproduction in Castor Bean

9

Meilian Tan, Lei Wang and Xingchu Yan

Abstract

Castor bean (*Ricinus communis* L.) is an important industrial oilseed crop with diverse sex patterns. Understanding the genetics and genomics of castor bean sexual reproduction is valuable for both its breeding and sex-determination researches. Here, we described the biology of flowering habit, sex forms, and sex variations in castor bean. Castor plant grows to raceme–cyme inflorescence and has the potential to be an ever-flowering perennial, it is normally monoecious raceme, but other inflorescence patterns such as pistillate, male, and intersperse also can be observed. A simplified overview on wide variation of sex tendency and variegation phenomenon in sex expression was given, and the factors affected sex expression were also summarized. Moreover, sex inheritance was also generalized, and the genes underlying or involved in the monocism, female, interspersed staminate flowers, and terminal hermaphroditism were mentioned. We reviewed the genetic variation of the replication for pistillate plant by the combination of tissue culture, molecular markers, and conventional breeding. It was showed that pistillate somaclones with high percentage of pistillate plants were obtained, and allelic differences were discovered among individuals of somaclonal populations with instable sex expression. Some genes associated with sex expression and reproduction, hormone response, transcription regulation and signal transduction, have also been identified by digital gene expression (DGE) analysis. This overview provided some information and valuable hints for insights into the molecular mechanism of sex determination.

9.1 Introduction

Sexual reproduction facilitates the long-term survival and adaptation of a species, and it is used for genetic improvement of economically important plants. Thus, basic understanding of the genetics and genomics of plant sexual reproduction will be valuable for both crop-breeding and sex-determination researches. Cross-pollination is a common reproduction mechanism in most species of flowering plants, and this type of sexual reproduction prevents the negative effects of inbreeding by promoting heterozygosity, genetic variability, and genetic exchange (Dellaporta and Calderon-Urrea 1993).

M. Tan · L. Wang · X. Yan (✉)
Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute (OCRI) of Chinese Academy of Agricultural Sciences, Wuhan 430062, People's Republic of China
e-mail: yanxc@oilcrops.cn

C. Kole and P. Rabinowicz (eds.), *The Castor Bean Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97280-0_9

Different modes of sexuality are observed in plants, and it can be classified at the level of individual flowers, individual plants, and plant populations (Dellaporta and Calderon-Urrea 1993).

Castor bean (*Ricinus communis* L.), as a versatile industrial oilseed crop, belongs to the Euphorbiaceae family with diverse sex patterns (Anjani 2012). Its sexuality changes occur occasionally in the individuals, especially in the female plants (Shifriss 1956). Hybrid breeding of castor bean for improved yield and oil purity is constrained by the genetic instability of female plants and the lack of in-depth knowledge of castor bean sexual reproduction (Tan et al. 2016). This chapter provides an introduction to sexual reproduction in castor bean from a genetics and genomics perspective.

9.2 The Biology of Sexual Reproduction in Castor Bean

9.2.1 Growth and Flowering Habit

Castor plant has the potential to be an ever-flowering perennial, but it can grow as an indeterminate annual or perennial depending on climate and soil types in the tropical, subtropical, and warm-temperate regions (Anjani 2012). After germination, the plant is composed by a main stem and all subsequently developed lateral shoots, which terminate in a floral cluster (Shifriss 1956). New shoots can grow from almost all the nodes, but occasionally, branching occurs only after the primary raceme ("raceme" represents "inflorescence", "primary raceme" means "the inflorescence that grows at the top of the main stem.") has appeared at the main apex. The number of nodes on the main stem varies widely with different varieties, from 6 to over 90 (Shifriss 1960), and which usually determine the plant height and bloom time. It was reported that a single individual of an early variety bearing over 200 inflorescences was found under favorable conditions (Shifriss 1956).

The inflorescence of castor bean plant is raceme–cyme and potentially ever-flowering. Usually, the plant of common varieties is monoecious bearing the racemose inflorescences (George and Shifriss 1967). The raceme is composed of the central axis and alternate branches bearing unisexual flowers. Staminate flowers are on the base and middle of raceme and female flowers on the upper part (Fig. 9.1a). Sex

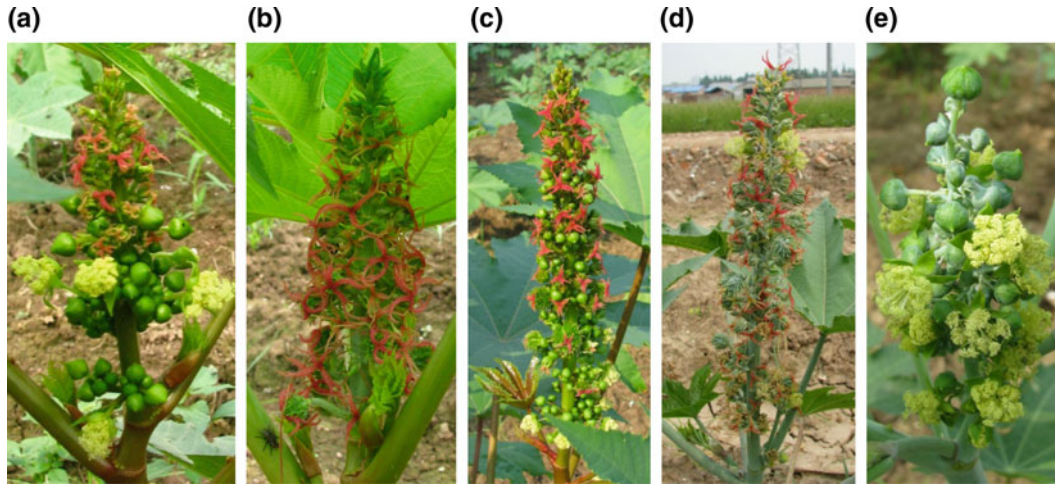

Fig. 9.1 Sexual pattern of racemes and sex forms in *Ricinus*. **a** Gradient monoecism (pistillate flowers on the apical distribution); **b** strictly pistillate inflorescence; **c** monoecious raceme with interspersed staminate flowers in the middle and apical region; **d** entirely interspersed (pistillate raceme with uniformly interspersed staminate flowers); **e** strictly male inflorescence (the flowers bear in the raceme are staminate flowers)

tendency of castor bean is represented by the ratio between pistillate and staminate flowers in the racemes. Higher or lower ratios can reveal variety's stronger female or stronger male tendencies, which depending on castor races (Shifriss 1956). Castor bean is wind-pollinated, but insect pollination by bees or flies has also been observed.

9.2.2 Sex Forms

The castor bean plant is normally monoecious (Fig. 9.1a), and wide diversity of inflorescence patterns can be observed in natural castor bean varieties. Its racemes include strictly pistillate (Fig. 9.1b, bearing only female flowers), male (Fig. 9.1e, only staminate flowers), apically interspersed (Fig. 9.1c, monoecism with interspersed male flowers in the apical pistillate region), and entirely interspersed (Fig. 9.1d, female and male flowers uniformly interspersed) (George and Shifriss 1967). In addition, inflorescence settings with one or a few hermaphrodite flowers occasionally occur (Jacob 1963). Correspondingly, individual castor bean plants show several sex modes, including normal monoecism, sex reversal, interspersed, and strictly female (Shifriss 1956; Jacob and Atsmon 1965; George and Shifriss 1967).

9.2.3 Sex Variation

Changes in sex expression occur in individuals of some plant species, including castor bean. Previous reports described a wide variation in sex tendency of inflorescences and individuals in castor bean, including strictly pistillate, sex-reversals and strictly male. The first report of a female castor bean plant was published in India in the 18th century (Roxburgh 1874), and wide variation in sex tendency of castor bean was reported much later (Joshi 1926; Katayama 1948). Shortly after, Shifriss (1956, 1960) and coworkers (George and Shifriss 1967) explained the variegation phenomenon in sex expression of castor bean. They found that sex-reversion was common in cultivated populations, and most of these changes were often attributed to momentary non-hereditary variations, although similar changes may be associated with hereditary instability (Shifriss 1956, 1960; George and Shifriss 1967). Here, we discussed the sex variation in castor bean and the factors that affect the sex expression.

Female racemes (especially primary pistillate racemes) and female individuals, can be found in castor bean fields, but most of them were proved to be spontaneous sex-reversals, because their following inflorescences were prone to revert into normal monoecious racemes. After investigating a mass of castor bean population in different areas, Shifriss (1956) found that the frequency of occurrence of sex-reversals among castor bean varieties ranged from 1:400 to 1:16,400, and the frequency is relatively higher in selected varieties and open-pollinated populations, than that in recently bred monoecious plant lines. He speculated that the causes of sex-reversion are both environmental and genetic (Shifriss 1956, 1961). However, as the majority of the observed spontaneous sex-reversals showed genetic instability, Shifriss considered that the occurrence of sex-reversion in castor bean is mostly due to genetic factors.

Sex expression of castor bean is mostly affected by light (Heslop-Harrison 1957), and other factors such as nutrient availability, temperature, humidity, developmental stage, as well as agricultural practices (e.g., sowing time and pruning), also play a role in sex tendency (Shifriss 1956). Female tendency is relatively strong in young plants (especially in primary racemes) under moderate temperature, moderate vegetative growth, and high concentrations of nutrients. Severe pruning of well-grown plants also promotes female tendency. In contrast, male tendency is stronger in old stocks or branch racemes that develop at later stages, and extreme temperatures, high vegetative growth and low concentrations of nutrients are favorable to male tendency. Furthermore, sex tendency also responds to seasonal variations, femaleness is strongest in spring and early summer, and strong male tendency is manifested during mid- and late

summer and winter for some castor bean varieties in Israel. Basing on sex tendency response to seasonal variations, a female inbred ("Nebraska") of castor bean was established, which showed strictly pistillate in spring and early summer.

As in other plant species, sex expression or tendency in castor bean is also affected by various plant growth regulators (Shifriss 1961; Tan et al. 2011, 2016). It has been reported that spraying a monoecious castor bean inbred with gibberellic acid (GA) can markedly increase female tendency (Shifriss 1961), auxin-like compounds was also shown to cause feminization of castor bean. The process of sex tendency can also be influenced by kinetin and morphactin (Kumar and Rao 1981), and it is shown that morphactin can increase male tendency in castor bean (Rkey 1978; Varkey and Nigam 1982), while daminozide can promote femaleness (Chauhan et al. 1987). It has also been shown that castor bean's flowering response to hormones may be different to other plants. For example, ethylene and ethylene-like substances (NIA 10637) result in maleness in castor bean, and can change female flowers into male ones in castor monoecious plants (Philipos and Narayanaswamy 1976), which is different from the observed response to these hormones in cucumber (Trebitsh et al. 1987; Tan et al. 2016).

9.3 Sex Inheritance in Castor Bean

In castor bean, gene *F* (for monoecism), control sex variants (genetically stable) ranging from female (*f*) to strongly male inbreds. Male tendency, sex-reversion, and sex instability could be affected by mutation of *F* or by an unknown factor that suppresses *F*. Monoecism is governed by qualitative genes that determine flower type, and genes that regulate gradient differentiation and sex tendency. In addition, gene modifiers such as *id* (*interspersed staminate flowers*) and *th* (*terminal hermaphroditism*) were also thought to affect sex differentiation. More evidences revealed that femaleness transmits more effectively to progeny, through female inflorescences of sex-reversal plants than that through monoecious racemes of the same plant (Shifriss 1956, 1960; Jacob and Atsmon 1965). The interspersed pattern of sex differentiation is determined by genes for femaleness and *id* genes, and the expression of interspersed staminate flowers depends on the dosage of two independent genes (*id*1 and *id*2), their loci, and the environment.

9.4 Genomics of Sex Expression in Castor Bean

Sex expression of castor bean has been described by previous research using traditional genetic and breeding means. However, we still know little about its molecular mechanisms, such as sex variation and genes determining sex expression in castor bean, including how to effectively conserve the pistillate individuals or lines.

In order to conserve or keep the pistillate lines for castor breeding and heterosis utilization, the pistillate individuals were replicated via a combination of tissue culture and conventional breeding. Sex expression of their F_1/F_2 progenies was observed, and somaclonal variation was detected by simple sequence repeat (SSR) markers (Tan et al. 2013). A total of 66 pistillate axillary buds of five genotypes were cultured on three kinds of media (initial culture, proliferation, and rooting), and 108 plantlets were formed and transplanted into the field with 50–70% survival. After investigating sex expression of the somaclones and their F_1/F_2 progenies, it was showed that the percentage of pistillate plants in the F_1 progeny obtained from female somaclones was higher (30.00–100.00%) than that in the progeny of its pistillate donors (26.67–50.00%) and the monoecious regenerations (0–5.00%). Two pistillate somaclones with high percentage of pistillate plants in the progeny populations (91.11 and 100.00% females in the F_1 progeny, and 76.10% and 75.00% females in the F_2 progeny from their sibling crosses, respectively) were obtained. The detection of somaclonal variation was performed using 14 genomic SSR and 19 EST-SSR (expressed sequence tag-simple sequence repeat) primer pairs. The results

showed that there was no allelic variation at SSR loci among the somaclones, which were derived from the same donor and with a stable pistillate trait. But allelic differences were discovered among individuals of somaclonal populations with instable sex expression.

In order to gain insights on sex expression and the molecular mechanisms of sex determination in castor bean, digital gene expression (DGE) analysis using the Illumina HiSeq 2000 was conducted to detect gene expression differences between apices and racemes for female (JXBM0705P) and monoecious (JXBM0705 M) lines, respectively. A total of 18 DGE libraries were constructed (apical buds at the 3–4-leaf stage, apical buds at the 5–7-leaf stage, and small racemes in 2–3 cm size, in triplicate), about six million tags for each library were produced, with more than 96% clear tags. Then, the differentially expressed genes (DEGs) were annotated by comparing with the reference genome of castor bean. It was shown that identical dynamic changes of gene expression existed in monoecious and female apical buds, during their transition from vegetative to reproductive development, with more genes expressing in the raceme formation and young racemes. The results showed that more than 3000 DEGs were identified, including a large number of DEGs potentially associated with sex expression and reproduction; other DEGs involved in hormone response and biosynthesis (such as auxin response and transport) were detected too. In addition, transcription regulation, signal transduction, histone demethylation/methylation, programmed cell death and pollination were also included. By comparing to the DEGs of the monoecism, it was suggested that many DEGs were suppressed at the 3–4-leaf stage in female line, but activated at the following stage (5–7-leaf), and finally, the number of upregulated DEGs was nearly even with those downregulated DEGs at the infant raceme stage.

So far, the molecular researches on sex expression are very poor and limited in castor bean, but these results provided some information and valuable hints for next insights into the molecular mechanism of sex determination.

References

Anjani K (2012) Castor genetic resources: a primary gene pool for exploitation. Indust Crop Prod 35:1–14

Chauhan SVS, Saxena BK, Kinoshita T (1987) Effect of daminozide (B9) on sex-expression and seed setting in castor bean, *Ricinus communis* L. Jpn J Breed 37: 262–266

Dellaporta SL, Calderon-Urrea A (1993) Sex determination in flowering plants. Plant Cell 5:1241–1251

George WL, Shifriss O (1967) Interspersed sexuality in *Ricinus*. Genetics 57:347–356

Heslop-Harrison J (1957) The experimental modification of sex expression in flowering plants. Biol Rev 32: 38–90

Jacob KM (1963) A trisomic male castor bean plant. J Hered 54:292–296

Jacob KM, Atsmon D (1965) Sex inheritance in *Ricinus communis* L.: evidence for a genetic change during the ontogeny of female sex reversals. Genetica 36:253–259

Joshi WV (1926) Some variations in the inflorescence of the castor plant. Poona Agri Coll Magaz 18:26–29

Katayama Y (1948) The progeny of female individuals in castor beans. Jpn J Genet 23:19 (in Japanese)

Kumar NR, Rao PG (1981) Activity of auxin like substances in relation to feminization of castor bean *(Ricinus communis* L.) influenced by kinetin and morphactin. Turrialba 31:201–208

Philipos H, Narayanaswamy S (1976) Induced potential for male sex expression in *Ricinus communis* L. by ethyl hydrogen-1-propylphosphonate. Protoplasma 87:71–77

Rkey VA (1978) M: experimental studies on effects of plant growth regulators on vegetative morphology and floral morphogenesis in *Ricinus communis* L. Ph.D. dissertation (Agronomy), Kanpur University, India

Roxburgh W (1874) Flora Indica. Thacker, Spink & Co, Calcutta, India

Shifriss O (1956) Sex instability in Ricinus. Genetics 41:265–280

Shifriss O (1960) Conventional and unconventional systems controlling sex variations in *Ricinus*. J Genet 57:361–388

Shifriss O (1961) Gibberellin as sex regulator in *Ricinus communis*. Science 133:2061–2062

Tan M, Yan M, Wang L, Yan X (2011) Effect of chemical treatments on sex expression of *Ricinus communis*. Chin Agri Sci Bull 27:164–169

Tan M, Yan M, Wang L, Yan X, Fu C, Wang L (2013) Replication of pistillate plants of *Ricinus communis* L. and investigation of the sex stability and genetic variation of the somaclones. Indust Crop Prod 50: 50–57

Tan M, Xue J, Wang L, Huang J, Fu C, Yan X (2016) Transcriptomic analysis for different sex types of *Ricinus communis* L. during development from apical buds to inflorescences by digital gene expression profiling. Front Plant Sci 6:1–16

Trebitsh T, Rudich J, Riov J (1987) Auxin, biosynthesis of ethylene and sex expression in cucumber (*Cucumis sativus*). Plant Growth Regul 5:105–113

Varkey M, Nigam RK (1982) A gradual reduction of female structures in the pistillate flowers of *Ricinus communis* L. (castor-bean) with chlorflurenol (morphactin). Biol Plant 24:152–154

Epigenetic Regulation in Castor Bean Seeds

10

Wei Xu and Aizhong Liu

Abstract

Endosperm genome hypomethylation is an evolutionarily conserved feature in flowering plants and plays important roles in the development of seeds by the control of gene expression and genomic imprinting. However, it is difficult to dissect the regulatory mechanisms of DNA methylation and genomic imprinting in most dicots, in which endosperm is ephemeral and disappears with seed development. Castor bean (*Ricinus communis*), unlike Arabidopsis (*Arabidopsis thaliana*), provides an excellent model for studying seed biology in dicotyledons due to its endosperm that is relatively large and persistent throughout the seed development. In this chapter, we dissect the potential regulatory mechanism of DNA methylation in castor bean seeds, emphasizing the processes of endosperm hypomethylation and CHH hypermethylation. We then identify and characterize the imprinted genes in castor bean endosperm. Their potential biological roles and epigenetic regulation were further analyzed and discussed.

W. Xu · A. Liu (✉)
Department of Economic Plants and Biotechnology, and Yunnan Key Laboratory for Wild Plant Resource, Kunming Institute of Botany, Chinese Academy of Sciences, 132 Lanhei Road, Kunming 650204, China
e-mail: liuaizhong@mail.kib.ac.cn

W. Xu
e-mail: xuwei@mail.kib.ac.cn

A. Liu
College of Forestry, Southwest Forestry University, Kunming 650224, China

10.1 DNA Methylation

10.1.1 The Regulatory Mechanisms of DNA Methylation in Plants

Epigenetic modifications are gene expression regulatory mechanisms that do not involve changes in the DNA sequence but can be inheritable. Currently, known epigenetic modifications fall within three types: DNA methylation, histone modification, and non-coding RNA (ncRNA)-associated gene silencing. Among them, DNA methylation has been well-studied and plays broad-ranging functional roles, including regulation of gene expression, chromatin organization, and suppression of mobile transposable elements (Heard and Disteche 2006; Klose and Bird 2006; Feinberg 2007).

Eukaryotic nuclear DNA can be methylated on carbon 5 of cytosine residues by enzymes called DNA methyltransferases and can be maintained through mitotic or meiosis cell division; thus, it is relatively stable and heritable epigenetic mark

C. Kole and P. Rabinowicz (eds.), *The Castor Bean Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97280-0_10

(Henderson and Jacobsen 2007). Although 5-methylcytosine (5mC) is the most common type of DNA methylation, studies have revealed that there are three other types of DNA methylation such as 4-methylcytosine (4mC), 5-hydroxy lmethylcytosine (5hmC), and 6-methyladenine (6mA). However, the levels and patterns of cytosine methylation appear to vary drastically among different organisms (Zhu 2009; Zemach et al. 2010a). In plants, 5mC generally occurs in three different nucleotide sequence contexts: symmetric CG and CHG sites and asymmetric CHH sites (where H = C, T or A). Generally, plant DNA methylation at symmetric CpG sites, where the cytosines are methylated on both DNA strands, is maintained upon DNA replication by the conserved DNA methyltransferase MET1 (METHYL TRANSFERASE 1) (Law and Jacobsen 2010). Symmetric CHG sites are methylated by the plant-specific DNA methyltransferase CMT3 (CHROMOMETHYLASE 3) and are dependent on histone H3 lysine 9 dimethylation (H3K9me2), which is established by the KRYPTONITE (KYP/SUVH4), SUVH5, and SUVH6 proteins (reviewed in Law and Jacobsen 2010). Plant asymmetric CHH sites are methylated de novo (all sequence contexts) and it is established by the RNA-directed DNA methylation (RdDM) pathway, which involves both small RNAs and long non-coding RNAs to recruit the DNA methyltransferase DRM1/2 (DOMAINS REARR ANGED METHYLTRANSFERASE 1 and 2) (Law and Jacobsen 2010; Mosher and Melnyk 2010). In addition, studies have demonstrated that CHH methylation could be produced by the plant-specific DNA methyltransferase CMT2 in an RdDM-independent pathway (Zemach et al. 2013; Stroud et al. 2014). Genome-wide plant DNA methylation analyses have revealed substantial amount of CG methylation and a small amount of non-CG methylation, although overall DNA methylation levels are broadly variable (Zemach et al. 2010a).

The extent and patterns of genomic DNA methylation are ultimately determined by the activity of DNA methyltransferases and demethylases. In Arabidopsis, active DNA demethylation mainly depends on the activity of the DNA glycosylase DEMETER (DME) (Kinoshita et al. 2004; Gehring et al. 2006), which can recognize and remove methylated cytosines by a base excision repair pathway (Law and Jacobsen 2010). Besides, other DNA glycosylases such as REPRESSOR OF SILENCING 1 (ROS1), DEM ETER-LIKE 2 (DML2), and DEMETER-LIKE 3 (DML3) have also been found to remove methylation from 5mC in Arabidopsis (Gong et al. 2002), protecting specific genomic regions from hypermethylation (Penterman et al. 2007). The activity and expression of these DNA methylation-related genes affect not only genomic DNA methylation levels, but also plant growth and development.

10.1.2 The Identification and Characterization of DNA Methylation-Related Genes in Castor Bean Seeds

By extensively searching the castor bean genome, Xu et al. (2016) identified and characterized eight genes encoding DNA methyltransferases and three genes encoding demethlylases. These DNA methylation proteins were classified as RcMET1-1 (29983.m003308) and RcMET1-2 (29609.m000606) in the MET group, RcCMT1 (28582.m000332) and RcCMT2 (29827. m002677) in the CMT group, RcDRM1(29631. m001043), RcDRM2 (29889.m003366), and RcDRM3 (29917.m001982) in the DRM group, and RcDnmt2 (29848.m004665) in the Dnmt2 group, according to their gene structures and protein domains (Fig. 10.1a). Similarly, three DNA demethylation proteins were identified as RcDME (29428.m000327), RcROS1 (29092. m000452), and RcDML3 (29991.m000647), respectively (Fig. 10.1b). Phylogenetic analyses indicated that each group of DNA methyltransferases displayed a monophyly with an evolutionary lineage from dicots to monocots (Fig. 10.1c), strongly implying that MET, CMT, DRM, and Dnmt2 might have independently evolved within angiosperms. As for DNA demethylase genes, phylogenetic analysis

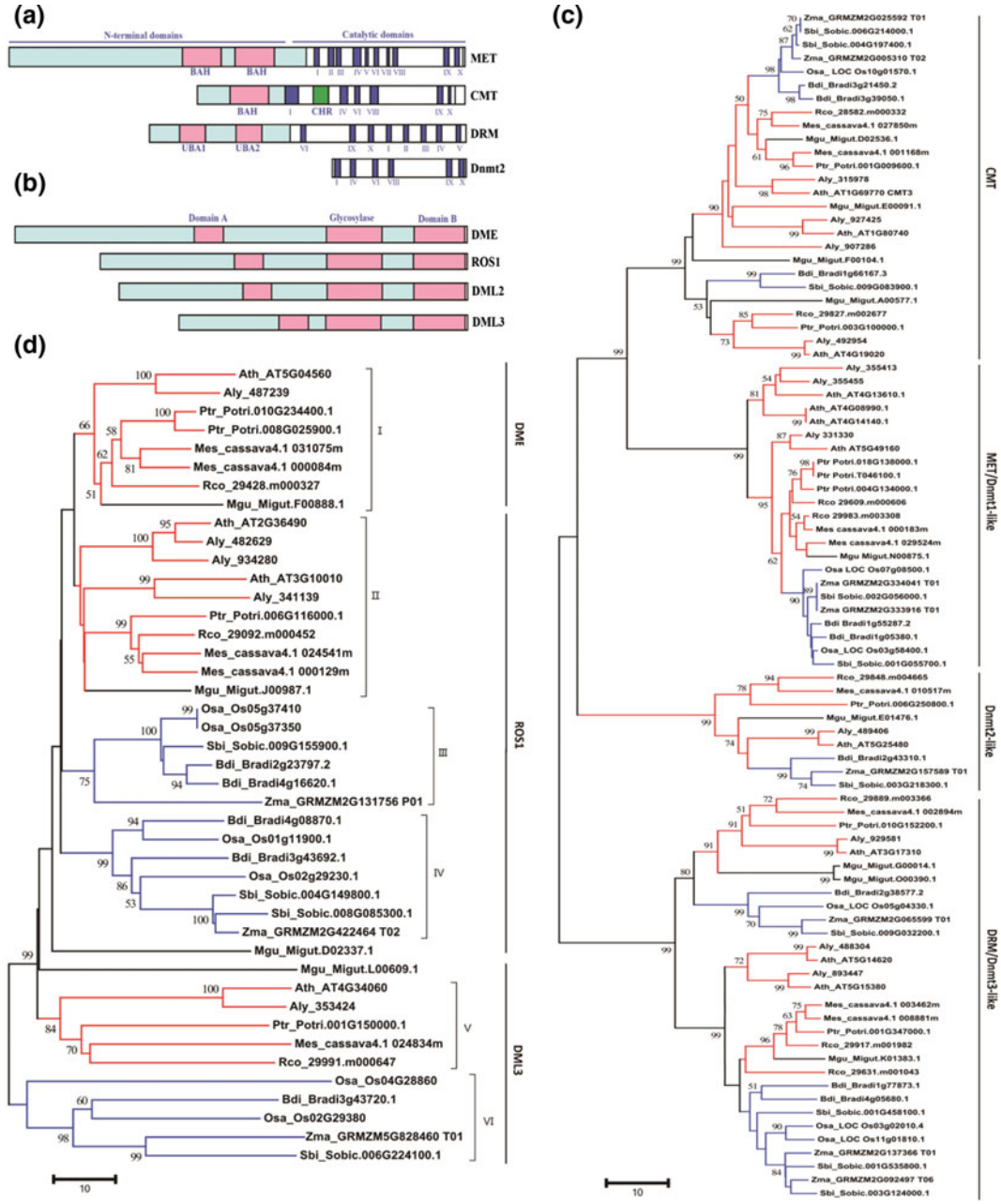

Fig. 10.1 Schematic structures of DNA methyltransferases and demethylases (**a** and **b**) and phylogenetic analysis (**c** and **d**). Monocot and dicot proteins are colored red and blue, respectively. Aly, *A. lyrata*; Ath, *A. thaliana*; Bdi, *Brachpodium distachyon*; Mes, *Manihot esculenta* (cassava); Mgu, *Mimulus guttatus* (monkey flower); Osa, *Oryza sativa* (rice); Ptr, *Populus trichocarpa* (poplar); Rco, *Ricinus communis* (castor bean); Sbi, *Sorghum bicolor*; Zma, *Zea mays* (maize) (Xu et al. 2016; www.plantphysiol.org; Reprinted with permission from Copyright American Society of Plant Biologists)

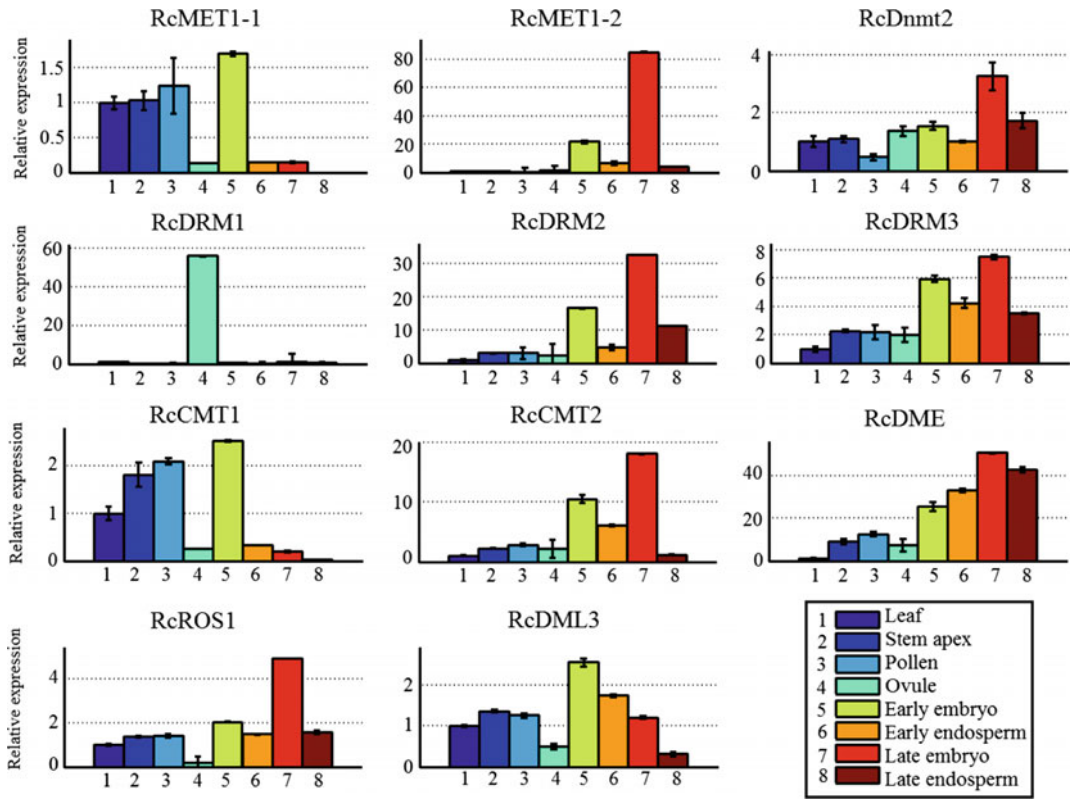

Fig. 10.2 Expression profiles of DNA methyltransferase and demethylase genes in different tissues of castor bean (Xu et al. 2016; www.plantphysiol.org; Reprinted with permission from Copyright American Society of Plant Biologists)

revealed that the DME orthologs were restricted to dicots only, including monkey flower (*Mimulus guttatus*), a basal dicot plant (Fig. 10.1d). This suggests that DME is probably monophyletic in dicots after a gene duplication occurred in early angiosperm evolution, while ROS1 and DML3 orthologs existed in both dicots and monocots. These observations strongly indicate that both methyltransferase and demethylase genes have evolved independently in angiosperms.

The expression profiles of the DNA methylation-related genes identified in castor bean were investigated in different tissues (Xu et al. 2016). Within the *MET1* group, *RcMET1-1* was highly expressed in leaf, stem apex, pollen, and early embryo, but lowly expressed in ovule and developing endosperm; while *RcMET1-2* was expressed only in developing embryos (particularly, at the late stage). Among the *CMT* genes, *RcCMT1* was highly expressed in leaf, stem apex, pollen, and early embryo, but lowly expressed in ovule and developing endosperm (including early and late stages); while the *RcCMT2* was expressed mainly in developing embryo tissues. Within the *DMR* genes, the *RcDRM1* was specifically expressed in ovule, while *RcDRM2* and *RcDRM3* were mainly expressed in developing embryo and endosperm tissues (Fig. 10.2).

Regarding the genes encoding DNA demethylases, the *RcDME* gene was broadly expressed in diverse tissues, particularly in embryo and endosperm tissues. This is distinct from the expression patterns of the DME gene observed in Arabidopsis, which was specifically expressed in the central cell of the female gametophyte or the vegetative cell of the male gametophyte (Choi et al. 2002; Schoft et al. 2011). Similarly, both *RcROS1* and *RcDML3*, which may be involved in mediating DNA demethylation, were extensively expressed in

diverse tissues (Fig. 10.2). In addition, it should be noted that the expression levels of most ***methylation-related*** genes were distinct between the early and late stages of endosperm and embryo development, implying that genomic DNA methylation levels within the endosperm and embryo might be variable through seed development.

10.1.3 Endosperm Hypomethylation in Castor Bean Seeds

Increasing evidence demonstrates that DNA methylation plays crucial roles in the control of seed development and seed size (Xiao et al. 2006a, b; Hu et al. 2014). Investigations on genomic DNA methylation in seeds of *Arabidopsis* (Gehring et al. 2009), rice (Zemach et al. 2010b), and maize (Zhang et al. 2014; Lu et al. 2015) revealed extensive genome hypomethylation in the endosperm, and it had a critical role in seed development. While comparing the genomic methylation levels between endosperm and embryo tissues of castor bean, Xu et al. (2014) found that the three types of DNA methylation were lower in the endosperm, showing 30.3, 18.3, and 11.2% methylation of CG, CHG, and CHH, respectively, while in the embryo the respective levels were 40.7, 24.0, and 12.7%. Compared with genomic DNA methylation levels in Arabidopsis (20.9% CG, 8.9% CHG, and 2.8% CHH in endosperm, and 26.9% CG, 10.6% CHG, and 4.4% CHH in embryo) (Hsieh et al. 2009), rice (36.0% CG, 9.7% CHG, and 0.65% CHH in endosperm, and 38.5% CG, 19.9% CHG, and 3.4% CHH in embryo) (Zemach et al. 2010b), and maize (75.4% CG, 55.7% CHG, and 1.2% CHH in endosperm, and 73.6% CG, 60.5% CHG, and 1.2% CHH in embryo) (Zhang et al. 2014), CHH methylation in castor bean endosperm and embryo were significantly higher. In castor bean seed, CHH methylation was predominant, accounting for about 68% of the total methylated cytosine, while CG and CHG methylation only accounted for about 16%, respectively, both in castor bean endosperm and embryo, distinct from that in Arabidopsis (Hsieh et al. 2009), rice (Zemach et al. 2010b), and maize (Zhang et al. 2014), in which CG methylation accounted for 45–55% of total methylated cytosine.

10.1.4 DNA Methylation and Gene Expression

Indeed, several studies also showed DNA hypomethylation could alter gene expression in the endosperm and embryo (Hsieh et al. 2009; Zemach et al. 2010b). By investigating the relationship between DNA methylation and expression of genes that were preferentially transcribed (at least fourfold higher) in endosperm relative to the embryo, Xu et al. (2016) found that genes that expressed preferentially in the endosperm showed significantly lower levels of CG and CHG methylation within the body of the gene and in flanking 2-kb regions compared to the embryo. As illustrated in Fig. 10.3, endosperm hypomethylation in some specific loci markedly promoted gene expression, whereas embryo hypermethylation near the promoter regions repressed gene transcription. For example, gene models annotated in the castor bean genome [28629.m000565, encoding a beta-fructofuranosidase involved in carbohydrate metabolism and 30093.m000370, homologous to AtYUC10 (AT1G48910) and involved in regulating seed development in Arabidopsis (Cheng et al. 2007)], exhibited higher expression levels and substantially lower DNA methylation in the endosperm than in the embryo. Furthermore, gene ontology (GO) function enrichment analysis (http://wego.genomics.org.cn/) of hypomethylated genes in endosperm predicted that these genes were substantially enriched in functions such as binding, catalytic, biological regulation, and cellular and metabolic processes (Fig. 10.4) (χ^2 test, $P < 0.05$). These data strongly imply that endosperm hypomethylation is likely involved in the regulation of endospermogenesis and biosynthesis of endosperm storage materials, though the molecular underpinnings are still unclear. In short, these results indicated that CG and CHG hypomethylation may be responsible for altering preferential gene expression in castor

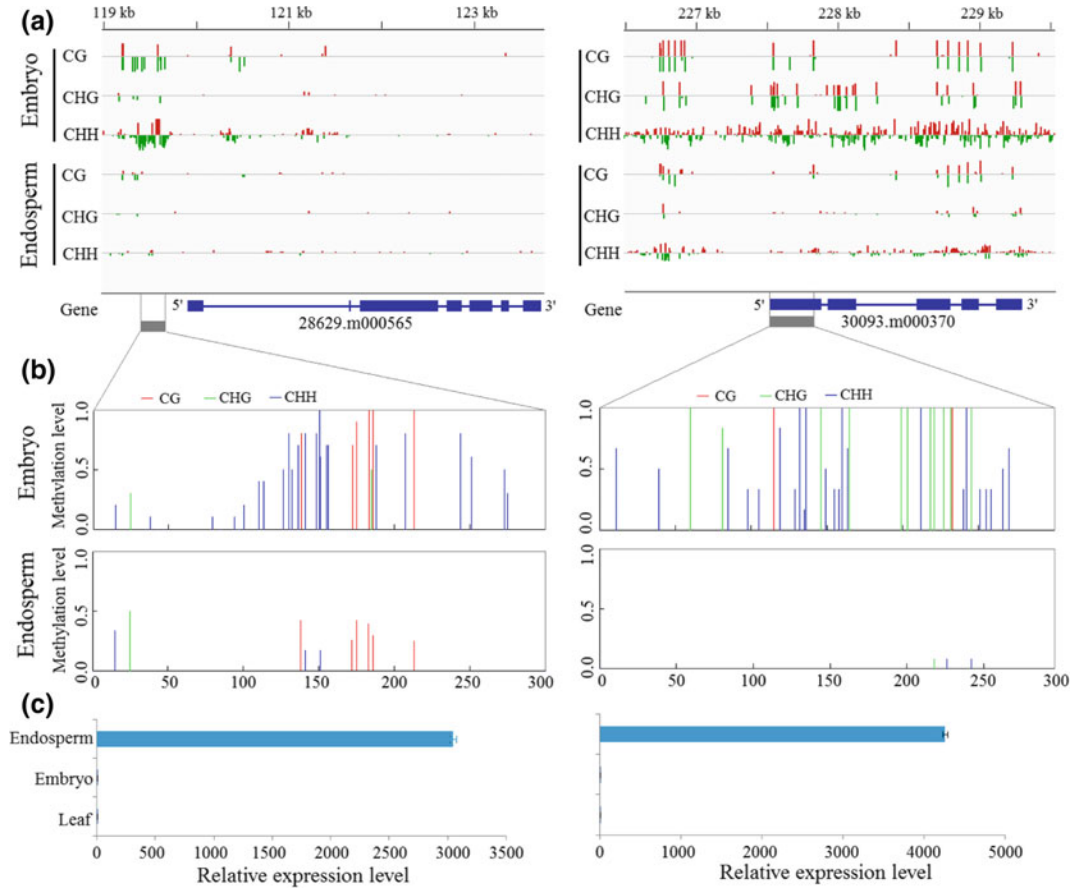

Fig. 10.3 Effect of two hypomethylated regions in endosperm gene expression. **a** DNA methylation level of all three sequence contexts in endosperm and embryo. The gray box indicates the region that was experimentally confirmed. **b** Validation of DNA methylation level by bisulfite PCR sequencing. **c** Relative expression level of two genes (28629.m000565 and 30093.m000370) overlapping the hypomethylation regions by qPCR analysis (Xu et al. 2016; www.plantphysiol.org; Reprinted with permission from Copyright American Society of Plant Biologists)

bean endosperm, particularly when CG and CHG hypomethylation occurs near or within a gene.

10.1.5 The Potential Regulatory Mechanism of DNA Methylation in Castor Bean Seeds

As mentioned above, extensive CG and CHG hypomethylation was observed in castor bean endosperm and high CHH methylation occurred in both endosperm and embryo. These distinct methylation patterns were conserved, implying that different mechanisms function to maintain DNA methylation in castor bean seeds. It is, however, unknown whether these patterns or mechanisms exist broadly within angiosperms since it has been studied in a limited number of species, including Arabidopsis, rice, and maize. In Arabidopsis, the *DMEs* were specifically expressed in the central cell of the female gametophyte,

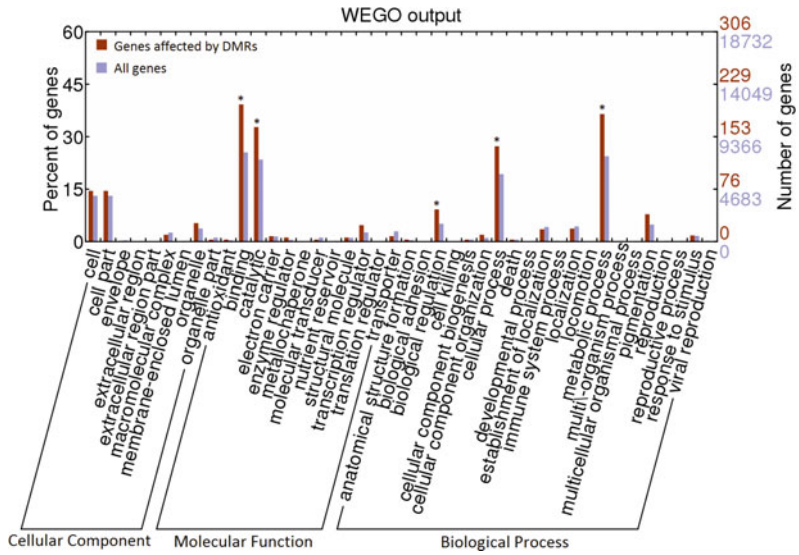

Fig. 10.4 GO enrichment analysis of genes with hypomethylation in endosperm. *indicates significant enrichment of GO terms (Xu et al. 2016; www.plantphysiol.org; Reprinted with permission from Copyright American Society of Plant Biologists)

where it probably plays a critical role in maintaining the genomic DNA methylation and resulting in DNA hypomethylation in the developing endosperm (Hsieh et al. 2009). Results from castor bean studies demonstrated that, as shown in Fig. 10.4, the *RcDMEs* were broadly expressed in various tissues, strongly suggesting that the mechanism that maintains genomic DNA methylation might be distinct from that in Arabidopsis. In addition, the expression of *RcMET1* and *RcCMT*, encoding CG and CHG methyltransferases, was substantially lower in the central cell of the castor bean female gametophyte or the developing endosperm relative to the embryo, implying that the activity of CG and CHG methyltransferases might be insufficient in the endosperm (which might result in a CG and CHG hypomethylation in endosperm). In particular, the expression of *RcDRM2* and *RcDRM3* was relatively higher in endosperm than in the embryo. These two proteins are probably responsible for maintaining genomic CHH methylation regulated by 24-nt small interfering RNAs (siRNAs) in the endosperm. This observation was supported by the expression profile of small RNAs in different tissues (Xu et al. 2013). The core components of the RdDM pathway, including *RcPoI IV* (30190.m011113), *RcPoI V* (30131.m007197), *RcRDR2* (30147.m013981) *RcDCL3* (29739.m003642), and *RcAGO4* (29684.m000322), were substantially up-regulated in endosperm and embryo tissues. These results suggest that genes involved in the RdDM pathway participate in altering non-CG methylation in the castor bean endosperm and embryo. Therefore, the RdDM pathway and RcDRMs in both embryo and endosperm may be critical and responsible for maintaining the non-CG methylation level between embryo and endosperm.

A proposed potential scenario is shown in Fig. 10.5 to explain the observed patterns of genomic DNA methylation in the embryo and endosperm of castor bean. The relatively low expression of *MET1* and *CMT* in the endosperm results in genomic CG and CHG hypomethylation due to insufficient activity of these enzymes, whereas the relatively higher expressions of *DRMs* (involving 24-nt siRNAs) lead to a maintenance of comparable CHH methylation levels in endosperm and embryo during endosperm development. Similarly, the relatively higher expression of *MET1*, *CMT,* and *DRM* genes in the embryo might maintain genomic CG, CHG, and CHH methylation during its development. Although this scenario just provides a preliminary hypothesis to explain a potential mechanism of genomic DNA methylation in castor bean seed, this studies added new insights to understand the molecular mechanism that drives this process.

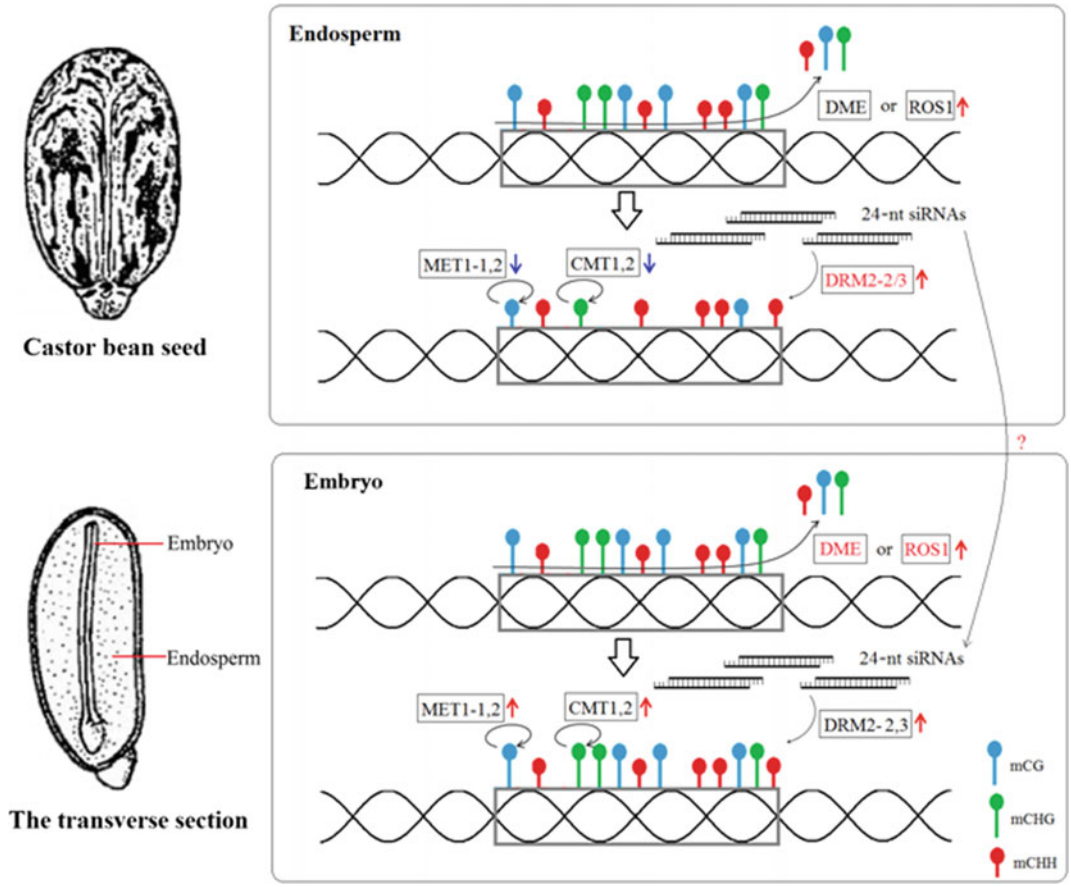

Fig. 10.5 A proposed model for explaining the DNA methylation pattern in castor bean endosperm and embryo. RcDME or RcROS1 removes the methylated cytosine at genome scale (blue lollipops for CG, green lollipops for CHG, and red lollipops for CHH methylation, respectively), and demethylation leads to the generation of 24-nt siRNAs that could guide de novo CHH DNA methylation by RcDRM2 in the endosperm and embryo; the insufficient endosperm expression of *RcMET1* and *RcCMT*, which maintain the CG and CHG methylation levels (denoted by looping arrow) result in reduced CG and CHG methylation (denoted by short-handled lollipops), compared with the embryo (denoted by long-handled lollipops) (Xu et al. 2016; www.plantphysiol.org; Reprinted with permission from Copyright American Society of Plant Biologists)

10.2 Genomic Imprinting

10.2.1 Genomic Imprinting and Its Potential Mechanism in Plants

Genomic imprinting is a common epigenetic mark that mainly arises in the plant endosperm, resulting in differential expression of maternal and paternal alleles, depending on their parental origin. The first discovery of an imprinted gene was the R-gene identified by a study of pigmentation of the outer layers of the endosperm in maize (Kermicle 1970, 1978), where the expression of the R allele depended on its parental origin, instead of on gene dosage. After that, approximately 20 new imprinted genes were identified and functionally characterized through genetic studies in some model plants such as Arabidopsis, rice, and maize (reviewed in Bauer and Fischer 2011). With the development of high-throughput sequencing technologies, genome-wide analyses have since led to the

identification of hundreds of imprinted genes in plants (see Table 10.1). Although genomic imprinting is largely confined to the endosperm tissue, gene imprinting has been found in embryo tissue (Jahnke and Scholten 2009; Waters et al. 2011; Nodine and Bartel 2012).

Typically, genomic imprinting leads to parent-of-origin specific differential expression of both maternally and paternally inherited alleles (Reik and Walter 2001; Feil and Berger 2007). In plants, genomic imprinting is predominantly found in endosperm tissues derived from the unique form of double-fertilization that occurs in flowering plants, wherein one sperm cell fertilizes a haploid egg cell, generating a diploid embryo, and another sperm cell fertilizes this diploid central cell, resulting in a triploid endosperm. While it is well-known that the function of the endosperm is to provide nutrition and ensure viability of the emerging plant after seed germination—that is why it constitutes up to 80% of a given seed—there are some intriguing biological aspects of gene imprinting that occurs only during endosperm development that remain largely unknown. The typical understanding of imprinting is that it arises as a by-product of a

Table 10.1 Genome-wide identification of imprinted genes in plants

Species	Reciprocal endosperms	Developmental stage[a]	MEGs[b]	PEGs[c]	References
Arabidopsis thaliana	Col × Bur	4 DAP	39	27	Wolff et al. (2011)
	Col × Ler	7–8 DAP	116	10	Hsieh et al. (2011)
	Col × Ler	6–7DAP	165	43	Gehring et al. (2011)
Capsella rubella	Cr A × Cr B	6 DAP	77	52	Hatorangan et al. (2016)
Ricinus communis	ZB107 × ZB306	35 DAP	200	9	Xu et al. (2014)
Zea mays	B73 × Mo17	14 DAP	54	46	Waters et al. (2011)
	B73 × Mo17	10 DAP	111	68	Zhang et al. (2011)
	B73 × Mo17	10–12 DAP	143	125	Zhang et al. (2014)
	B73 × Mo17	7 DAP	36	80	Xin et al. (2013)
	B73 × Mo17	10 DAP	184	50	Xin et al. (2013)
	B73 × Mo17	15 DAP	15	48	Xin et al. (2013)
	B73 × Mo17	14 DAP	31	56	Waters et al. (2013)
	B73 × Ki11	14 DAP	28	55	Waters et al. (2013)
	Mo17 × Ki11	14 DAP	9	24	Waters et al. (2013)
	B73 × Oh43	14 DAP	39	76	Waters et al. 2013
	Mo17 × Oh43	14 DAP	25	45	Waters et al. (2013)
Oryza sativa	Nipponbare × 93-11	5–6 DAP	93	72	Luo et al. (2011)
Sorghum bicolor	BTx623 × Y49	14 DAP	85	16	Zhang et al. (2016)
Triticum aestivum	Y177 × RM220 (diploid)	15–20 DAP	62	29	Yang et al. (2018)
	Jinying8 × SCAUP (tetraploid)	15–20 DAP	90	45	Yang et al. (2018)
	Doumai and Keyi5214 (hexaploid)	15–25 DAP	94	52	Yang et al. (2018)

[a]DAP: days after pollination
[b]MEG: maternally expressed genes
[c]PEG: paternally expressed genes

silencing mechanism targeting invading foreign DNA (Gehring et al. 2009). The cause of gene imprinting has been attributed to a parental conflict over the distribution of resources from mother to offspring (Haig 2004). This parental conflict hypothesis argues that maternally expressed genes would evenly distribute resources among progeny to maximize seed set, while paternally expressed genes would promote the growth of individual progeny (Haig 2013). There is also evidence that imprinted genes affect demand and supply of nutrients in both the mammalian placenta and during plant endosperm development (Reik et al. 2003).

Both DNA methylation and chromatin modifications have been widely proposed as the major underlying molecular mechanisms that modulate the expression of imprinted genes. It has been demonstrated that the establishment of DNA methylation marks differentially in plant parental genomes mainly relies on the activity of the DNA glycosylase DME (Choi et al. 2002). During gametophyte development, DME is specifically expressed in the central cell of the female gametophyte, and demethylates maternal alleles prior to fertilization, whereas the paternal alleles remain methylated owing to the lack of DME activity in sperm cells (Choi et al. 2002; Kinoshita et al. 2004; Jullien et al. 2006). This results in differential DNA methylation levels between maternal and paternal genomes and the establishment of genomic imprinting in *Arabidospsis*. Alongside the suggestive evidence of the role of DNA methylation in gene imprinting, there is also evidence that chromatin modifications mediated by polycomb group proteins (PRC2) contribute to gene regulation by imprinting (Köhler and Weinhofer 2010; Raissig et al. 2011). DME-mediated DNA demethylation in the central cell results in the expression of maternally imprinted genes such as MEDEA (or MEA) (Grossniklaus et al. 1998; Kinoshita et al. 1999), FIS2 (Jullien et al. 2006), FIE1 (Hermon et al. 2007), which are core components of the FIS-PRC2 complex. FIS-PRC2 catalyzes the trimethylation of histone H3 on lysine 27 (H3K27me3)—an important repressive epigenetic mark—of maternal genome in the central cell of the female gametophyte, resulting in the differential expression of alleles in a parent-of-origin manner.

10.2.2 Genomic Imprinting in the Castor Bean Endosperm

Recent whole-genome surveys have now revealed abundant imprinted genes in plants, but the evolutionary and biological processes of imprinted genes in plants remain to be fully understood. Insights into this important question are likely to come from the identification and comparison of imprinted genes across different species, especially among dicotyledonous (dicots) plants. It is relatively difficult to identify imprinted genes in most dicots including Arabidopsis because cellularized endosperms do not undergo significant proliferative growth and are instead absorbed by the developing embryo during seed development (Sreenivasulu and Wobus 2013). By contrast, in some dicot species (such as castor bean, physic nuts, and coffee), the endosperm undergoes an intense period of cell division immediately after cellularization, followed by differentiation and specialization that result in the formation of a nutritive tissue for supporting the embryo and the seedling during germination (Forbis et al. 2002; Sreenivasulu and Wobus 2013).

Given the state of the research on genomic imprinting among dicots, identifying and separating imprinted genes in the endosperms would be greatly beneficial for dissecting the genetic mechanism underlying this phenomenon in plants. Castor bean provides an excellent system for identifying imprinted loci and characterizes parent-of-origin effects of imprinted genes in endosperm development among dicots, because the castor bean endosperm is relatively large and persists throughout seed development (Greenwood and Bewley 1982) and it can be easily

separated from the embryo, avoiding DNA contamination between embryo and endosperm during sampling. More importantly, the castor bean genome has been well-annotated, providing a great deal of information on genomic organization and transposable elements (Chan et al. 2010).

Xu et al. (2014) performed genome resequencing and identified allelic SNPs between two elite castor bean inbred lines ZB306 (with small seed size) and ZB107 (with large seed size). Using ZB306 and ZB107 as parents, they constructed two mRNAs libraries of developing endosperm from reciprocal crosses and conducted global transcription analyses. According to the differential expression of alleles in reciprocal seeds, they identified a total of 209 imprinted genes, including 200 maternally expressed genes (MEGs) and nine paternally expressed genes (PEGs). Further, 85% of selected 67 imprinted genes were validated using RT-PCR sequencing. Interestingly, while tracking the imprinting status of several loci during endosperm development, they found that most of the imprinted genes exhibited a dynamic expression pattern during endosperm development. That said, it is important to emphasize that this list of imprinted genes is specific to very discrete stages of the castor bean life-cycle (i.e., endosperm development) and the number of imprinted genes in castor bean could potentially be larger if imprinting also occurs in mature tissues. By extension, it is quite possible that the expression of imprinted genes during endosperm development may be affected by reprogramming of epigenetic marks following fertilization. Likewise, some imprinted genes were identified to be accession-dependent. This type of gene imprinting may be an epigenetic natural variation (Vaughn et al. 2007), and one worth further inquiry.

10.2.3 Potential Biological Roles and Conservation Among Species

At 35 days after pollination (DAP), the cellular endosperm increases dramatically in volume, leading a rapid increase of seed weight, similar to what occurs during the IV–V stage of seed development (Greenwood and Bewley 1982). Based on GO function enrichment analysis, imprinted genes identified in this stage were significantly enriched in the GO terms cell, cell part, binding, catalytic, and cellular process and metabolic process. A majority of MEGs were involved in metabolic process, suggesting these genes are likely associated with specific endospermogenesis processes.

Furthermore, a comparison of the 209 imprinted genes from castor bean with the known imprinted genes of Arabidopsis (Gehring et al. 2011; Hsieh et al. 2011; Wolff et al. 2011), rice (Luo et al. 2011) and maize (Waters et al. 2011; Zhang et al. 2011; Xin et al. 2013) showed that only 8 imprinted genes were conserved between Arabidopsis and castor bean, 5 between rice and castor bean, and 12 between maize and castor bean. Low conservation of imprinted genes has also been observed between different seed developmental stages or among different accessions within a species (Hsieh et al. 2011; Xin et al. 2013; Waters et al. 2013). The substantial variation in the targets of imprinting in different species implies the existence of different mechanisms for the initiation and maintenance of imprinted expression patterns in plants (Waters et al. 2013). Transposable elements (TEs) are extensively methylated in somatic tissues but become demethylated during seed development. Usually, the number and types of TEs vary largely among different species. To test whether imprinted genes were always associated with the

TEs, Xu and his colleagues investigated the number and types of TEs within 4 kb regions flanking imprinted and non-imprinted genes (Xu et al. 2014). The result showed that specific TEs (LTR/gypsy family) were enriched around MEGs ($P = 3.46^{-e}$) in castor bean when compared with non-imprinted genes, which is markedly different from that observed in Arabidopsis (Wolff et al. 2011) and Maize (Zhang et al. 2011). Moreover, several imprinted genes located in syntenic regions between Arabidopsis and castor bean are surrounded by totally different sets of TEs. Therefore, TEs are considered a major driving force for gene imprinting (Gehring et al. 2009; Hsieh et al. 2009; Wolff et al. 2011).

Overall, these results showed that imprinted genes are involved in a variety of functions related to regulating endosperm development. The limited conservation of imprinted genes between castor bean and other studied plant species indicated that the mechanisms underlying the beginning and maintenance of gene imprinting could be widely variable in different plants. Importantly, these findings suggest that the differences in gene imprinting between dicots and monocots may not have resulted from their separate evolutionary history. This is at best a logical supposition or extension, and further investigation into the functional differentiation of imprinted genes identified in different plants will likely provide much greater insight into understanding the mechanisms behind gene imprinting in plants.

10.2.4 Regulation of DNA Methylation on Imprinted Genes

DNA methylation has long been regarded as a key player in epigenetic regulation. Differential methylation levels of the maternal and paternal alleles in endosperm tissues have been identified in Arabidopsis (Gehring et al. 2009; Hsieh et al. 2009) rice (Zemach et al. 2010) and maize (Lauria et al. 2004). Genome-wide DNA methylation patterns can be analyzed by bisulfite sequencing (BS-seq). BS-seq analysis of embryo and endosperm tissues of castor bean showed genome-wide hypomethylation of the endosperm, as compared to the embryo. Moreover, a few MEGs exhibited a dramatic differential methylation levels between the two parental alleles. The maternal hypomethylation of MEGs observed in castor bean endosperm implies that there could potentially be a DNA glycosylase DME gene in castor bean (i.e., *RcDME,* gene ID: 29428.m000327 in the castor bean genome annotation) functioning in the central cell that causes hypomethylation of the endosperm, though the function of the predicted *RcDME* gene remains to be determined. More likely, the repression of *RcMET1* gene (a DNA methyltransferase) in the central cell and/or the endosperm is critical for the establishment of differential DNA methylation between parental alleles (Xu et al. 2016). It should be noted that most of the imprinted genes identified in castor bean display no differential methylation level between the parental alleles, suggesting that DNA methylation may have a limited effect on the establishment of gene imprinting and other factors (such as histone modifications) may the causal mechanisms underlying the rise of gene imprinting in castor bean. Such possibilities remain to be investigated, but it is possible that partially imprinted genes might be expressed in the seed coat and transported into the endosperm embedded in developmental maternal seed coat tissues.

References

Bauer MJ, Fischer RL (2011) Genome demethylation and imprinting in the endosperm. Curr Opin Plant Biol 14:162–167

Chan AP, Crabtree J, Zhao Q et al (2010) Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotechnol 28:951–956

Cheng Y, Dai X, Zhao Y (2007) Auxin synthesized by the YUCCA flavin monooxygenases is essential for

embryogenesis and leaf formation in Arabidopsis. Plant Cell 19:2430–2439

Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, Jacobsen SE, Fischer RL (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. Cell 110:33–42

Feil R, Berger F (2007) Convergent evolution of genomic imprinting in plants and mammals. Trends Genet 23:192–199

Feinberg AP (2007) Phenotypic plasticity and the epigenetics of human disease. Nature 447:433–440

Forbis TA, Floyd SK, de Queiroz A (2002) The evolution of embryo size in angiosperms and other seed plants: implications for the evolution of seed dormancy. Evolution 56:2112–2125

Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, Goldberg RB, Fischer RL (2006) DEMETER DNA glycosylase establishes MEDEA Polycomb gene self-imprinting by allele-specific demethylation. Cell 124:495–506

Gehring M, Bubb KL, Henikoff S (2009) Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science 324:1447–1451

Gehring M, Missirian V, Henikoff S (2011) Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. PLoS ONE 6:e23687

Gong Z, Morales-Ruiz T, Ariza RR, Roldán-Arjona T, David L, Zhu JK (2002) ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. Cell 111:803–814

Greenwood JS, Bewley JD (1982) Seed development in *Ricinus communis* castor bean. I. Descriptive morphology. Can J Bot 60:1751–1760

Grossniklaus U, Vielle-Calzada JP, Hoeppner MA, Gagliano WB (1998) Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. Science 280:446–450

Haig D (2004) Genomic imprinting and kinship: how good is the evidence? Annu Rev Genet 38:553–585

Haig D (2013) Kin conflict in seed development: an interdependent but fractious collective. Annu Rev Cell Dev Biol 29:189–211

Hatorangan MR, Laenen B, Steige K, Slotte T, Köhler C (2016) Rapid evolution of genomic imprinting in two species of the Brassicaceae. Plant Cell 28:1815–1827

Heard E, Disteche CM (2006) Dosage compensation in mammals: fine-tuning the expression of the X chromosome. Gene Dev 20:1848–1867

Henderson IR, Jacobsen SE (2007) Epigenetic inheritance in plants. Nature 447:418–424

Hermon P, Srilunchang KO, Zou J, Dresselhaus T, Danilevskaya ON (2007) Activation of the imprinted Polycomb Group Fie1 gene in maize endosperm requires demethylation of thematernal allele. Plant Mol Biol 64:387–395

Hsieh TF, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, Zilberman D (2009) Genome-wide demethylation of Arabidopsis endosperm. Science 324:1451–1454

Hsieh TF, Shin J, Uzawa R, Silva P, Cohen S, Bauer MJ, Hashimoto M, Kirkbride RC, Harada JJ, Zilberman D, Fischer RL (2011) Regulation of imprinted gene expression in Arabidopsis endosperm. Proc Natl Acad Sci USA 108:1755–1762

Hu L, Li N, Xu C, Zhong S, Lin X, Yang J, Zhou T, Yuliang A, Wu Y, Chen YR et al (2014) Mutation of a major CG methylase in rice causes genome-wide hypomethylation, dysregulated genome expression, and seedling lethality. Proc Natl Acad Sci USA 111:10642–10647

Jahnke S, Scholten S (2009) Epigenetic resetting of a gene imprinted in plant embryos. Curr Biol 19:1677–1681

Jullien PE, Kinoshita T, Ohad N, Berger F (2006) Maintenance of DNA methylation during the Arabidopsis life cycle is essential for parental imprinting. Plant Cell 18:1360–1372

Kermicle JL (1970) Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. Genetics 66:69–85

Kermicle JL (1978) Imprinting of gene action in maize endosperm. In: Walden DB (ed) Maize Breeding and Genetics. Wiley and Sons, Hoboken, NJ, USA, pp 357–371

Kinoshita T, Yadegari R, Harada JJ, Goldberg RB, Fischer RL (1999) Imprinting of the MEDEA polycomb gene in the Arabidopsis endosperm. Plant Cell 11:1945–1952

Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X et al (2004) One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. Science 303:521–523

Klose RJ, Bird AP (2006) Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci 31:89–97

Köhler C, Weinhofer I (2010) Mechanisms and evolution of genomic imprinting in plants. Heredity 105:57–63

Lauria M, Rupe M, Guo M, Kranz E, Pirona R, Viotti A, Lund G (2004) Extensive maternal DNA hypomethylation in the endosperm of *Zea mays*. Plant cell 16:510–522

Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11:204–220

Lu X, Wang W, Ren W, Chai Z, Guo W, Chen R, Wang L, Zhao J, Lang Z, Fan Y, Zhao J, Zhang C (2015) Genome-wide epigenetic regulation of gene transcription in maize seeds. PLoS ONE 10: e0139582

Luo M, Taylor JM, Spriggs A, Zhang H, Wu X, Russell S, Singh M, Koltunow A (2011) A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. PLoS Genet 7: e1002125

Mosher RA, Melnyk CW (2010) siRNAs and DNA methylation: seedy epigenetics. Trends Plant Sci 15:204–210

Nodine MD, Bartel DP (2012) Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. Nature 482:94–97

Penterman J, Zilberman D, Huh JH, Ballinger T, Henikoff S, Fischer RL (2007) DNA demethylation in the Arabidopsis genome. Proc Natl Acad Sci USA 104:6752–6757

Raissig MT, Baroux C, Grossniklaus U (2011) Regulation and flexibility of genomic imprinting during seed development. Plant Cell 23:16–26

Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. Nat Rev Genet 2:21–32

Reik W, Constancia M, Fowden A, Anderson N, Dean W, Ferguson-Smith A, Tycko B, Sibley C (2003) Regulation of supply and demand for maternal nutrients in mammals by imprinted genes. J Physiol 547:35–44

Schoft VK, Chumak N, Choi Y, Hannon M, Garcia-Aguilar M, Machlicova A, Slusarz L, Mosiolek M, Park JS, Park GT et al (2011) Function of the DEMETER DNA glycosylase in the Arabidopsis thaliana male gametophyte. Proc Natl Acad Sci USA 108:8042–8047

Sreenivasulu N, Wobus U (2013) Seed-development programs: a systems biology-based comparison between dicots and monocots. Annu Rev Plant Biol 64:189–217

Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE (2014) Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat Struct Mol Biol 21:64–72

Vaughn MW, Tanurdzić M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A et al (2007) Epigenetic natural variation in *Arabidopsis thaliana*. PLoS Biol 5:e174

Waters AJ, Makarevitch I, Eichten SR, Swanson-Wagner RA, Yeh CT, Xu W, Schnable PS, Vaughn MW, Gehring M, Springer NM (2011) Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. Plant Cell 23:4221–4233

Waters AJ, Bilinski P, Eichten SR, Vaughn MW, Ross-Ibarra J, Gehring M, Springer NM (2013) Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. Proc Natl Acad Sci USA 110:19639–19644

Wolff P, Weinhofer I, Seguin J, Roszak P, Beisel C, Donoghue MT, Spillane C, Nordborg M, Rehmsmeier M, Köhler C (2011) High-resolution analysis of parent-of-origin allelic expression in the Arabidopsis endosperm. PLoS Genet 7:e1002126

Xiao W, Brown RC, Lemmon BE, Harada JJ, Goldberg RB, Fischer RL (2006a) Regulation of seed size by Hypomethylation of maternal and paternal genomes. Plant Physiol 142:1160–1168

Xiao W, Custard KD, Brown RC, Lemmon BE, Harada JJ, Goldberg RB, Fischer RL (2006b) DNA methylation is critical for Arabidopsis embryogenesis and seed viability. Plant Cell 18:805–814

Xin M, Yang R, Li G, Chen H, Laurie J, Ma C, Wang D, Yao Y, Larkins BA, Sun Q et al (2013) Dynamic expression of imprinted genes associates with maternally controlled nutrient allocation during maize endosperm development. Plant Cell 25:3212–3227

Xu W, Cui Q, Li F, Liu A (2013) Transcriptome-wide identification and characterization of microRNAs from castor bean *Ricinus communis* L. PLoS ONE 8:e69995

Xu W, Dai MY, Li F, Liu AZ (2014) Genomic imprinting, methylation and parent-of-origin effects in reciprocal hybrid endosperm of castor bean. Nucl Acids Res 42:6987–6998

Xu W, Yang T, Dong X, Li DZ, Liu A (2016) Genomic DNA methylation analyses reveal the distinct profiles in castor bean seeds with persistent endosperms. Plant Physiol 171:1242–1258

Yang G, Liu Z, Gao L, Yu K, Feng M, Yao Y, Peng H, Hu Z, Sun Q, Ni Z, Xin M (2018) Genomic imprinting was evolutionarily conserved during wheat polyploidization. Plant Cell 30:37–47

Zemach A, Kim MY, Silva P, Rodrigues JA, Dotson B, Brooks MD, Zilberman D (2010a) Local DNA hypomethylation activates genes in rice endosperm. Proc Natl Acad Sci USA 107:18729–18734

Zemach A, McDaniel IE, Silva P, Zilberman D (2010b) Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 14:916–919

Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D (2013) The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell 153:193–205

Zhang M, Zhao H, Xie S, Chen J, Xu Y, Wang K, Zhao H, Guan H, Hu X, Jiao Y et al (2011) Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. Proc Natl Acad Sci USA 108:20042–20047

Zhang M, Xie S, Dong X, Zhao X, Zeng B, Chen J, Li H, Yang W, Zhao H, Wang G et al (2014) Genome-wide high resolution parental-specific DNA and histone methylation maps uncover patterns of imprinting regulation in maize. Genome Res 224:167–176

Zhang M, Li N, He W, Zhang H, Yang W, Liu B (2016a) Genome-wide screen of genes imprinted in sorghum

endosperm, and the roles of allelic differential cytosine methylation. Plant J 85:424–436

Zhang Y, Mulpuri Sujatha, Liu A (2016b) Photosynthetic responses of seed coat under high light exposure increases lipid accumulation in non-green seeds of castor (*Ricimus communis* L.). Photosynth Res 128 (2):125–140

Zhu JK (2009) Active DNA demethylation mediated by DNA glycosylases. Annu Rev Genet 43:143–166

Oil Biosynthesis and Biotechnology in the Castor Bean

11

Federico García Maroto and Diego López Alonso

Abstract

The castor bean plant, *Ricinus communis* L., is an oleaginous species that produces a very unusual oil, composed mainly of ricinoleic acid (ca. 90%). This fatty acid is a hydroxylated derivative (18:1OH) of common oleic acid, and it has a great range of industrial applications. As a result, castor oil is regarded as one of the most important renewable resources for the chemical industry. However, there are certain drawbacks in the use of this plant due to the presence of highly toxic ricin in its seeds, as well as some strong allergens found both in seed and pollen. Current knowledge of the metabolic routes involved in castor oil biosynthesis, and the molecular mechanisms responsible for the efficient channeling of ricinoleic acid into the seed oil, have opened the way for genetic engineering approaches aimed at producing castor-like oils from common oleaginous species. Domestication of the castor plant through biotechnological strategies is also being envisaged as a feasible alternative to improve the common source of ricinoleic acid. This chapter describes the current knowledge of oil biosynthesis in castor bean, as well as biotechnological approaches being explored to improve the available sources of ricinoleic acid.

Abbreviations

CoA	Coenzyme A
CPT	Choline phosphotransferase
DGAT	Diacylglycerol acyltransferase
DGTA	Diacylglycerol:diacylglycerol transacylase
ER	Endoplasmic reticulum
FA	Fatty acid
FAD2	Fatty acid desaturase 2
FAH12	Fatty acid hydroxylase 12
G3P	Glycerol-3-phosphate
GPAT	Glycerol-3-phosphate acyltransferase
LACS	Long-chain acyl-CoA synthetase
LPA	Lysophosphatidic acid
LPAT	Lysophosphatidate acyltransferase
LPC	Lysophosphatidylcholine
LPCAT	Lysophosphatidylcholine acyltransferase
MBOAT	Membrane Bound O-acyltransferase
OA	Oleic acid
Ole-CoA	Oleoyl-CoA
Ole-PC	sn-2-oleoyl-phosphatidylcholine
PAH	Phosphatidic acid hydrolase
PC	Phosphatidylcholine
PDAT	Phospholipid:diacylglycerol acyltransferase
PLA2	Phospholipase A2
PLC	Phospholipase C
RA	Ricinoleic acid

F. García Maroto (✉) · D. López Alonso
Grupo de Investigación "Biotecnología de Productos Naturales" (BIO-279), CEIA3, BITAL, Universidad de Almería, CITEIIB, 04120 Almería, Spain
e-mail: fgmaroto@ual.es

C. Kole and P. Rabinowicz (eds.), *The Castor Bean Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97280-0_11

Ric-CoA	Ricinoleoyl-CoA
TAG	Triacylglycerol
TFA	Total fatty acids

11.1 Introduction

Castor bean (*Ricinus communis* L.) is an oleaginous plant that is well known around the world, mainly for the traditional use of its oil as a laxative. It is believed that castor oil was first used as an ointment 4000 years ago in Egypt, from where it spread to other parts of the world, including Greece and Rome, 2500 years ago (Chan et al. 2010). Nevertheless, castor oil has some special properties that still makes it useful nowadays for a wide range of industrial applications.

The oil content of castor bean seed ranges from 35 to 65% (50% average). This variation is largely due to genetic differences between cultivars (Yadav and Anjani 2017). Castor oil is composed of up to 90% ricinoleic acid (RA, 18:1OH), an oleic acid derivative (OA, 18:1) hydroxylated in the 12-carbon that bears interesting chemical properties. This unique fatty acid is the source of many chemical derivatives currently used in industrial applications (McKeon 2016; Vinay et al. 2016).

There are four abundant triacylglycerol (TAG) species that constitute castor oil: triricinolein, 1,2-diricinoleoyl-3-oleoyl-sn-glycerol, 1,2-diricinoleoyl-3-linoleoyl-sn-glycerol, and 1,2-diricinoleoyl-3-linolenoyl-sn-glycerol (Lin et al. 1998). In addition to its high concentration in castor oil, RA is biosynthesized at technical grade purity allowing for its direct industrial use (McKeon et al. 1999; McKeon 2016).

The world production of castor oil is far less than that of major plant oils such as palm, soybean, and rapeseed; the yearly production of each being around 60 million ton (data from FAOSTAT http://www.fao.org/faostat). However, there is growing industrial interest in castor oil as indicated by the rise in castor oil trade. Indeed, in the period from 2004 to 2013, the number of tons traded in castor oil doubled (Fig. 11.1).

Nevertheless, castor bean poses certain environmental and health concerns that impedes its large-scale development as a useful biotechnological plant. The main issue is the presence of ricin, an extremely toxic protein, deadly even in minute quantities, which poisons the residues produced after oil extraction. Although these residues are usually treated using high temperature to denature the proteins, there is a concern about the effectiveness of this process to reach a complete inactivation of ricin.

Another problem of castor bean relates to the presence of several powerful allergens in the seed, which are responsible for multiple

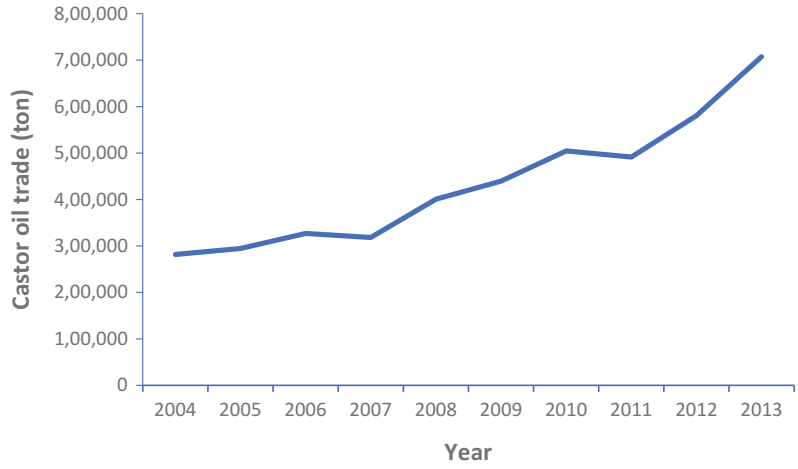

Fig. 11.1 Evolution of castor oil production along the 2004–2013 period. Data taken from FAOSTAT http://www.fao.org/faostat

secondary health problems in people exposed to its cultivation, harvesting, and processing.

Moreover, castor bean is not well adapted to temperate climates, it is not amenable to common agronomic practices, and the whole process from cultivation to seed oil extraction is very labor-intensive. However, the existence of a current market demand for castor oil is generating various attempts to improve the agronomic characteristics of this plant and/or transfer castor bean's oil-producing capacity to a crop that is free from the above-mentioned problems.

Any attempt to biotechnologically produce castor-like oil will require in-depth knowledge of oil biosynthesis in castor bean. In the following sections, we provide a review of castor oil biosynthesis, focusing on the roles played by the enzymes and genes involved in relevant metabolic routes. Current efforts to develop castor oil biotechnology are also considered.

11.2 Oil Synthesis in Castor Bean

The view of plant oil biosynthesis as a linear metabolic route (the Kennedy pathway) comprising the progressive acylation of glycerol-3-phosphate has significantly changed over the last 60 years. Current understanding of plant oil biosynthesis comprises a complex metabolic network of multiple pathways located in different subcellular compartments, fed by a variety of metabolite pools that lead to the production of TAGs or membrane glycerolipids (Bates 2016).

Although TAGs are synthesized in different cellular compartments (the chloroplast and the endoplasmic reticulum), and in specific tissues or organs, the seed is usually the specialized organ for oil accumulation, where massive TAG synthesis takes place. Castor bean accumulates 50–60% of its oil in its seeds. Therefore, we will focus on TAG biosynthesis in the endoplasmic reticulum of the developing castor bean seed endosperm.

De novo fatty acid biosynthesis takes place in the chloroplast, where the most abundant fatty acids are produced and then transferred to the cytosol (Bates et al. 2009) following activation as coenzyme A (CoA) thioesters (acyl-CoA) in the outer plastid membrane. Activated fatty acids are the precursors of TAGs and membrane lipids.

All the enzymes involved in TAG biosynthesis are located in the endoplasmic reticulum as integral membrane proteins. In addition, they are probably forming several multi-enzymatic complexes by establishing direct contact with their partners, thus facilitating coordinated activity.

11.2.1 Fatty Acids Modification and Acyl Editing

Fatty acids (FAs) synthesized in the plastid and transferred to the initial acyl-CoA pool are mainly 18:1 (oleic acid, OA), 16:0 (palmitic acid), and 18:0 (stearic acid). However, membrane and storage (oil) lipids are also composed of other modified FAs, such as 18:2 (linoleic acid), and 22:1 (eicosenoic acid). Therefore, these modified FAs must be previously synthesized through specific reactions to be ready for incorporation into different glycerolipids. It is well known that, in plants, most of these modifications (e.g., desaturation, epoxidation, and hydroxylation) take place on the acyl groups esterified to phosphatidylcholine (PC) at the *sn*-2 central position of glycerol. After being modified, acyls can return to the acyl-CoA pool (Bates et al. 2012; Haslam et al. 2016). This process, usually called acyl editing, is catalyzed by the coordinated action of lysophosphatidylcholine acyltransferase (LPCAT) and either phospholipase A_2 (PLA2) or reverse LPCAT activities. LPCAT transfers the acyl group to lysophosphatidylcholine (LPC) giving PC an acyl at the *sn*-2 ready to be edited. Next, the edited acyl can return to the acyl-CoA pool by the sequential action of PLA2 and the long-chain acyl-CoA synthetase (LACS) or, alternatively, by the reverse reaction catalyzed by the LPCAT (Stymne and Stobart 1984) (Fig. 11.2). As will be described in the next section, RA synthesis in castor bean is just a variation of this general plant process. As will be described, plants also have additional mechanisms allowing the incorporation of the edited acyls into glycerolipids.

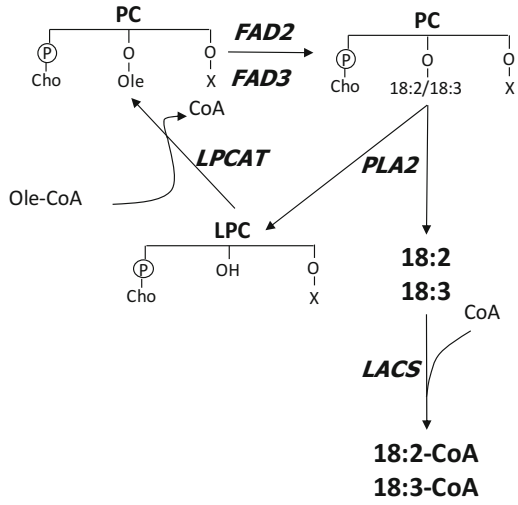

Fig. 11.2 Acyl-editing mechanism illustrating the incorporation of linoleic (18:2) and α-linolenic (18:3) acids into the acyl-CoA pool after being synthesized by desaturases FAD2 and FAD3, using oleic acid attached to PC as a substrate

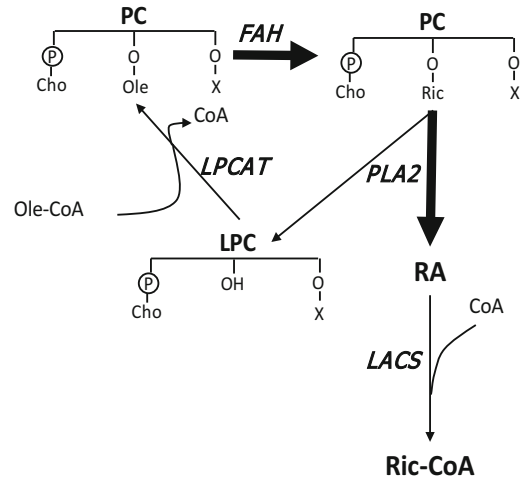

Fig. 11.3 Acyl-editing mechanism showing the incorporation of ricinoleic acid into the acyl-CoA pool after being synthesized by the hydroxylase FAH12, using as a substrate oleic acid attached to PC

11.2.2 Ricinoleic Acid Biosynthesis

The biosynthesis of RA has been relatively well known for some time (Moreau and Stumpf 1981). The reactions take place exclusively in the endoplasmic reticulum of seed cells, using *sn*-2-oleoyl-phosphatidylcholine (Ole-PC) as the substrate. The process starts when oleoyl-CoA (Ole-CoA) is efficiently transferred to the *sn*-2 position of lysophosphatidylcholine (LPC) by the LPCAT, rendering the PC with the OA esterified at *sn*-2 (Fig. 11.3). It is in this form that the OA is hydroxylated in the twelfth carbon by oleoyl-12-hydroxylase, commonly known as fatty acid hydroxylase 12 (FAH12). RA is subsequently released from the PC by PLA2 activity and immediately activated with CoA by LACS, rendering the ricinoleoyl-CoA (Ric-CoA) into the acyl-CoA pool, which is ready for use in TAG synthesis (Fig. 11.3).

Recent studies (Lager et al. 2013) have given further support to the old proposal (Stymne and Stobart 1984) that the reverse reaction of LPCAT can significantly contribute to the acyl-editing mechanism. Hence, experiments performed with castor LPCAT have demonstrated that RA release from PC, as a result of this reverse reaction, is much more efficient than for OA (Lager et al. 2013). This indicates that LPCAT activity could also contribute to the flux of RA into TAG synthesis although its extension, relative to the PLA2 mechanism, is still to be determined.

It should be noted that the release of RA mediated by PLA2 also produces LPC, which is re-esterified with new Ole-CoA by the LPCAT, thus closing the RA biosynthesis cycle (Fig. 11.3). This is recognizable as a particular castor bean adaptation of the general acyl-editing process, also known as the "Lands cycle". Therefore, the biosynthesis of Ric-CoA, the activated form of RA, mainly involves four enzymes: LPCAT, FAH12, PLA2, and LACS.

In castor bean, over 70% of the seed TAGs contain three RA (triricinolein) (McKeon et al. 1999). This requires ~90% of seed OA to flow through PC for conversion into RA (Bates and Browse 2011). The LPCAT activity is a key in supplying the PC substrate for hydroxylation (McKeon 2016). Brown et al. (2012) detected the transcription of two *LPCAT* candidate genes in castor bean seed, but neither showed differential expression between leaf and seed. They found another candidate in a protein from the

membrane bound O-acyltransferase (MBOAT) super-family, which showed raised transcription in the seed. This gene was later cloned (Arroyo-Caro et al. 2013a), and it was shown that its enzymatic product preferentially uses monounsaturated FAs (e.g., OA); this was in agreement to previous experiments with microsomal extracts where an enzymatic activity responsible for incorporating oleate (from Ole-CoA) into the *sn*-2 position of lyso-PC was detected (McKeon et al. 1997). Transcription analysis of this gene revealed generalized expression in diverse organs, as well as a higher transcript level in the developing seed, consistent with the transcriptomic data from Brown et al. (2012). However, constant levels of the transcript were recorded throughout seed development, thus indicating that basal LPCAT expression in the seed is sufficient to support Ole-PC requirements for RA synthesis (Arroyo-Caro et al. 2013a).

A single castor bean FAH12 gene (Brown et al. 2011) was identified and cloned based on the assumption that the enzyme responsible for the hydroxylation reaction generating RA must have evolved from fatty acid desaturase 2 (FAD2), thus maintaining high sequence homology (van de Loo et al. 1995). FAH12 is only expressed in seed endosperm (van de Loo et al. 1995; Brown et al. 2012), and its expression pattern and enzymatic timing activity were correlated with TAG accumulation in the developing castor bean seed (Brown et al. 2012). These features of FAH12, which catalyzes one of the key regulatory steps for the synthesis of castor oil (McKeon et al. 1999; Lu et al. 2007), make it an essential gene for any biotechnological strategy related to RA production.

The release of RA by PLA2 seems to be very specific. Two genes coding for this enzyme have been cloned in castor bean—one of them showed 60% greater release activity for RA than for other fatty acids. The protein encoded by this gene has homology to an Arabidopsis lipase (GLIP) belonging to the GDSL family, but its specific involvement in the release of the hydroxylated fatty acid (HFA) for TAG biosynthesis has still to be demonstrated (Brown et al. 2012). The released RA must be esterified to CoA by LACS activity. Up to seven putative-*LACS* genes have been detected in castor bean based on transcriptional analysis. One of them, an ortholog of the Arabidopsis *LACS9* gene, showed higher activity in the castor bean endosperm (Brown et al. 2011, 2012). Among the seven isoforms, a gene-designated *RcACS2* (putatively orthologous to Arabidopsis *LACS6*) has been shown to encode an enzyme with a preference for RA as a substrate (He et al. 2007). This gene could be a valuable tool for biotechnological purposes even though it seems to correspond to a peroxisomal isoenzyme and accordingly, it is predominantly expressed in the germinating seed.

11.3 TAG Biosynthesis in Castor Bean

Initial studies on the castor plant suggested that TAG was synthesized through a linear mechanism, the well-known Kennedy pathway or glycerol-3-phosphate pathway following the sequential direct incorporation of three acyls to the glycerol moiety with an additional reaction to hydrolyse the phosphate from *sn*-3 in the phosphatidic acid (Bafor et al. 1991) (Fig. 11.4). Research results over the last ten-to-twenty years have shown that the actual pathway is best represented as a network of different reactions. Nevertheless, the Kennedy pathway has a central position in the biosynthesis of TAG and membrane lipids, where additional reactions either depart from a Kennedy pathway intermediary or feeds into the Kennedy pathway (Fig. 11.4).

11.3.1 Lysophosphatidate Synthesis

The initial reaction of the Kennedy pathway is the incorporation of the first acyl group to the *sn*-1 position of glycerol-3-phosphate (G3P) that produces lysophosphatidic acid (LPA). This reaction is catalyzed by the glycerol-3-phosphate acyltransferase (GPAT) enzyme (Fig. 11.4). Although a plant microsomal GPAT has been searched for over many years, the identity of the

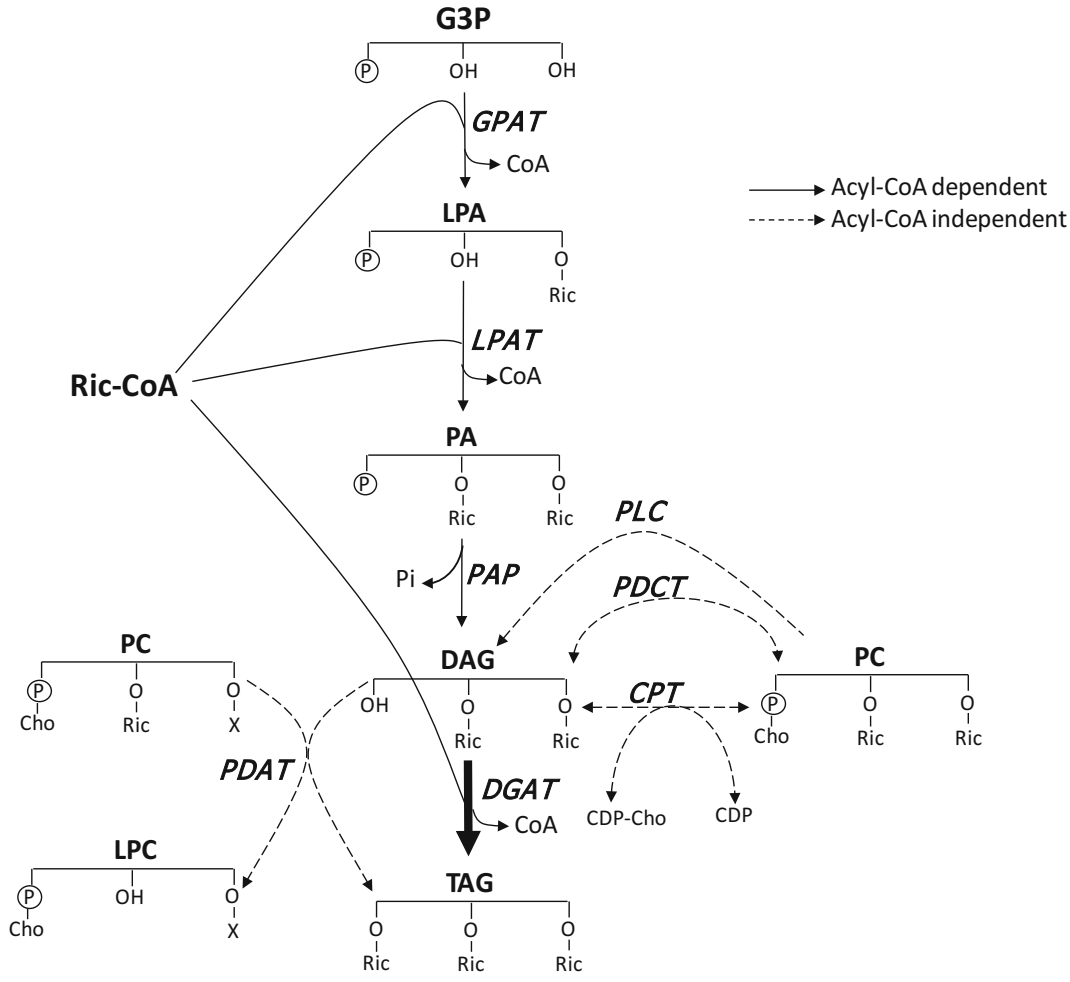

Fig. 11.4 Diverse pathways for the synthesis of triacylglycerol in plant seeds. See the text for further explanation

gene encoding this enzyme has remained elusive until recently. The GPAT9 of *Arabidopsis* has been shown to be directly involved in TAG biosynthesis (Shockey et al. 2016) and, in fact, it is an *sn*-1 acyltransferase with a high preference for acyl-CoA (Singer et al. 2016). Interestingly, the expression pattern of this gene in *Arabidopsis* seems to be constitutive (Singer et al. 2016).

Searches of the castor bean genome sequence failed to identify a GPAT candidate with a specific role in seed TAG synthesis. This is probably because the predicted GPAT expression level in the seed was relatively low in the two developing endosperm stages checked (Brown et al. 2012). Brown et al. (2012) speculated that a castor bean GPAT9 ortholog appeared to be the main candidate involved in castor oil synthesis as it was the most abundant among the GPATs transcripts found in the endosperm, although similar expression levels were evident in other tissues, as was the case in *Arabidopsis*. However, characterization of the GPAT9 castor enzyme has not yet been reported.

11.3.2 Synthesis of Phosphatidate

The lysophosphatidic acid acyltransferase (LPAAT) enzyme, also called lysophosphatidate acyltransferase (LPAT), catalyzes the second

acylation step, at the *sn*-2 position in glycerol, to yield phosphatidic acid (PA) (Fig. 11.4). PA is a key metabolite per se (e.g., as a signal molecule), and it is also a branching point, leading to the synthesis of diverse phospholipids in one direction and to DAG/TAG synthesis in the other. Therefore, PA metabolism should be under tight regulatory control.

In plants, LPAT shows high substrate selectivity, and microsomal LPATs, in particular, usually have preference for certain acyl groups in plant species that accumulate unusual fatty acids in their TAGs (Frentzen 1998; Millar et al. 2000; Snyder et al. 2009). Two classes of microsomal LPAT genes (class A and class B) have been described (Frentzen 1998; Arroyo-Caro et al. 2013b) within a very large family of related sequences (LPAT-like acyltransferases containing the plsC domain). Class A genes are ubiquitously present in different plant organs and tissues and have a broad (unspecific) acyl preference. Class B genes, evolutionarily more related to prokaryotic LPATs, are typically expressed in seed. Class B LPATs also exhibit a preference for the unusual and specific fatty acids produced in their plant species, as exemplified by the *Limnanthes* (Brown et al. 2002) and *Cocos* (Knutzon et al. 1995) LPAT enzymes which are more active on substrates containing erucic and lauric acid, respectively. The absence of B-class genes in Arabidopsis (Kim et al. 2005) seems to indicate that this type of LPAT may not be essential, at least in certain plant species.

Initial studies of LPAT activity in microsomal preparations from the castor bean showed selective enzymatic activity on Ole-CoA and linoleoyl-CoA while, intriguingly, Ric-CoA was not incorporated into PA unless polyamines were present (Tomosugi et al. 2006). However, our group later cloned two *RcLPAT* genes that are expressed in seed (each belonging to one of the two gene classes—A or B) and their products were shown to be active without requiring polyamines (Arroyo-Caro et al. 2013b). The class A gene, designated *RcLPAT2*, is highly expressed in castor bean seed and at a higher rate than the class B gene *RcLPATB*. Complementation experiments carried out with a defective *plsC* mutant from *Escherichia coli* and biochemical analyses demonstrated that both genes encoded enzymes with LPAT activity, although with very different acyl-CoA preferences (Arroyo-Caro et al. 2013b).

Enzymatic assays on RcLPAT2 indicated a preference for Ric-CoA over other fatty acid thioesters when ricinoleoyl-LPA is used as the acyl acceptor, while Ole-CoA is the preferred substrate when oleoyl-LPA is employed (Arroyo-Caro et al. 2013b). Therefore, the action of the first acyltransferase of the Kennedy pathway, RcGPAT, determines the output from the LPAT reaction. If RA is introduced by GPAT, then RcLPAT2 will favor the entrance of a second RA molecule. This constitutes a channeling mechanism that would favor the synthesis of RA-enriched TAGs in the castor bean seed. As observed for orthologs from other plants, *RcLPAT2* is not only expressed in the castor bean seed but also in other organs where RA is not synthesized. In this case, the specificity of RcLPAT2 would also allow the synthesis of OA-containing PA as a precursor for phospholipid synthesis. Consequently, the characteristics of microsomal GPAT activity in different castor bean tissues may be a key to understanding TAG synthesis in castor bean, which will be facilitated by further knowledge on the *Ricinus* microsomal GPAT.

The properties of RcLPAT2 make it a useful target for metabolic engineering, as has been recently shown (Chen et al. 2016b). In that work, an increase in RA at the *sn*-2 position of TAGs was obtained by expressing *RcLPAT2* in *Lesquerella*.

RcLPATB is also expressed in different castor bean organs, and it shows a preference for saturated acyls such as 12:0, 14:0, and 16:0 versus monounsaturated fatty acids, regardless of the LPA acceptor. This feature was unexpected given the known activities shown by analogous class B enzymes from *Limnanthes* and *Cocos*, which preferred the unusual acyls particular to each species (22:1 and 12:0, respectively). Interestingly, up-regulation of *RcLPATB* was observed during the oil-accumulation phase of seed development. However, given its acyl

preference, its role in lipid synthesis remains obscure (Arroyo-Caro et al. 2013b).

Other extra-plastidial LPAT isoenzymes have been identified in *Ricinus*, such as RcLPAT3A and RcLPAT3B, but they are not expressed in the seed (Arroyo-Caro et al. 2013b).

11.3.3 Diacylglycerol Synthesis

The next step in the Kennedy pathway is the hydrolysis of the PA phosphate group to produce DAG. This reaction is catalyzed by a phosphatidic acid phosphatase (PAP) (Fig. 11.4) also called phosphatidic acid hydrolase (PAH). As PA, DAG plays additional roles in lipid-signaling cascades, besides being a precursor of certain phospholipids. Therefore, the existence of a tight control over DAG synthesis is also predicted (Bates et al. 2009).

Eleven *PAP* candidate genes have been identified in Arabidopsis, probably with different specificities and biochemical functions. Two of them, *AtPAH1* and *AtPAH2*, have high homology with well-characterized *PAH* genes from yeast, but seed specificity or a role in oil biosynthesis has not been demonstrated for either of the two genes. A double-knockout mutant of *AtPAH1* and *AtPAH2* showed impairment in the phospholipid turnover but TAG synthesis remained unaffected (Li-Beisson et al. 2013). It is likely that some of the 11 candidate genes are involved in DAG synthesis, but which of them is currently unknown.

Not surprisingly, PAP is represented in castor bean by a multigene family. However, there is no evidence that any of them is involved in TAG synthesis (Brown et al. 2012).

11.3.4 Alternative DAG Sources

Besides the Kennedy pathway, several other DAG metabolic routes have been described. Several reactions that allow interconversion between DAG and PC are known and some of those have been linked to oilseed metabolism (Bates et al. 2009). The reaction catalyzed by the enzyme CDP-choline:diacylclyerol cholinephosphotransferase (CPT), which is responsible for the synthesis of PC from DAG, has been known for a long time. The reversibility of this reaction allows the conversion of PC into DAG. More recently, a phosphatidyl-choline:diacylglycerol choline-phospho-transferase activity (PDCT; also known as Reduced Oleate Desaturation 1, ROD1) has been described. PDCT catalyzes the reversible transformation of PC into DAG by transferring the phosphocholine group between both molecules (Lu et al. 2009; Bates 2016; Haslam et al. 2016). Phospholipase C (PLC) can also generate DAG from PC by removing the phosphocholine group. These enzymes can generate a PC-derived DAG pool, different from the de novo DAG pool produced through the Kennedy pathway, which could be used for TAG synthesis.

The PC-derived DAG pool has been shown to play a major role in TAG synthesis in some plant species. In soybeans, over 95% of TAG was estimated to come from the PC-derived DAG pool (Bates et al. 2009), and Arabidopsis can utilize PC-derived DAG to produce as much as 93% of the total TAG synthesis (Bates and Browse 2011).

In vitro experiments using *Ricinus* endosperm microsomes showed that a substantial part of TAG synthesis is carried out using the "de novo" DAG pool (Bafor et al. 1991). However, there are also indications that some "de novo" DAG containing RA may be used for PC synthesis, which in turn, could generate diricinoleoyl DAG derived from PC, through the activity of some of the above-mentioned enzymes, such as PDCT (Bates and Browse 2012).

A castor bean PDCT gene (*RcROD1*), homologous to the *ROD1* gene from Arabidopsis, has been cloned (Van Erp et al. 2011; Hu et al. 2012), and it was able to slightly increase RA accumulation (ca. 5%) when expressed in transgenic Arabidopsis plants cotransformed with FAH12 (Van Erp et al. 2011). This suggests that PC-derived DAG might be used to some extent for TAG synthesis in *Ricinus* (Bates and Browse 2012).

There is no available information the activities of CPT and PLC, which could potentially generate DAG from PC.

11.3.5 Diacylglycerol Acylation

11.3.5.1 Acyl-CoA-Dependent TAG Synthesis

The last step of the Kennedy pathway, namely TAG synthesis by a third DAG acylation using acyl-CoA as a donor (acyl-CoA-dependent synthesis), is catalyzed by diacylglycerol acyltransferase (DGAT; also called DAGAT). DGAT adds the third and last acyl to the *sn*-3 glycerol position producing TAG (Fig. 11.4). In castor bean, this activity has proven to be very specific to Ric-CoA (He et al. 2006; Burgal et al. 2008). Two evolutionarily unrelated DGAT genes, *RcDGAT1* and *RcDGAT2*, have been found in castor bean. A predicted ortholog of the peanut soluble DGAT has also been identified in the castor bean genome (Chan et al. 2010; Brown et al. 2012).

The relative importance of the role played by both microsomal DGAT enzymes (1 and 2) in TAG synthesis in plants has been a matter of controversy (Shockey et al. 2006; Chen et al. 2007; Mañas-Fernández et al. 2009; Arroyo-Caro et al. 2016). It has been proposed that DGAT2 plays a principal role, particularly in plants that produce rare oils, such as *Vernonia*, *Lesquerella*, and *Ricinus*, in which DGAT2 would specialize in channeling these unusual fatty acids (RA in the case of *Ricinus*) toward TAG, avoiding their incorporation into membrane lipids (Shockey et al. 2006). DGAT1 would be less specific regarding both expression and preferred acyl substrate. Some experimental results provide partial support for this hypothesis in *Stokesia laevis*, *Lesquerella*, and *Vernonia* (Yu et al. 2006). However, experimental evidence in *Ricinus* indicate a somewhat different situation as described below (Chen et al. 2007; Burgal et al. 2008).

RcDGAT1 was the first DGAT gene to be cloned and characterized in castor bean. It showed much higher activity with 1,2-diricino leoyl-glycerol than with 1,2-dipalmitoyl-glycerol or 1,2-dioleoyl-glycerol (He et al. 2004b; McKeon and He 2015). A study of its regulation in developing castor bean seed showed a pronounced delay in protein appearance relative to transcription of its corresponding mRNA. Nonetheless, RcDGAT1 protein levels correlated fairly well with DGAT activity and TAG accumulation. Consequently, the authors predicted that DGAT1 is the predominant DGAT activity in developing castor bean seed (He et al. 2004a).

RcDGAT2 was later cloned and the encoded protein was shown to be very specific for both substrates, Ric-CoA and 1,2-diricinoleoyl-glycerol, making up the typical castor oil triricinolein. Moreover, when *RcDGAT2* was co-expressed with *FAH12* in Arabidopsis, it dramatically boosted the HFA content from about 18–28%, whereas the *RcDGAT1* gene did not (Burgal et al. 2008). RcDGAT2 is currently considered the main enzyme responsible for DGAT activity in seed, that results in high RA content in castor oil (Kroon et al. 2006; Brown et al. 2012; McKeon 2016), despite being considerably less active than RcDGAT1 (McKeon 2016). Nevertheless, it has been conclusively proven that RcDGAT1 also play a part in TAG synthesis in castor bean seed (McKeon and He 2015). Further studies comparing *RcDGAT1* expression profiles among different castor bean tissues reported a higher abundance of *RcDGAT1* mRNA in leaf than in seed, casting doubt about which DGAT genes are more relevant for TAG synthesis and at what developmental stage (Chen et al. 2007).

11.3.5.2 Acyl-CoA-Independent TAG Synthesis

Various enzymatic activities have been described that acylate the DAG molecule to yield TAG, in which the acyl donor is not acyl-CoA. Those mechanisms of TAG synthesis are designated as acyl-CoA-independent. The most common acyl-CoA-independent activity employs a phospholipid molecule (usually PC) as the donor, and catalyzes the transfer of the fatty acid from the *sn*-2 position of PC to the *sn*-3 position of DAG. This enzyme, designated phospholipid:diacylglycerol acyltransferase [PDAT; (Dahlqvist et al. 2000)] was initially described a few years ago in *Saccharomyces cerevisiae*. Soon afterward, orthologs were found in a variety of plant species (Fig. 11.4).

In castor bean, PDAT activity was eight times higher with RA than with OA in the *sn*-2 position

of PC (McKeon 2016). Up to seven PDAT genes have been reported in castor bean. Two of them, *PDAT1A* and *PDAT2*, exhibited seed-specific expression, indicating their involvement in TAG biosynthesis. This suggestion is supported by the fact that *PDAT1A* expression increases the proportion of HFA in oils from transgenic plants (van Erp et al. 2011).

It is well known that in some plants the acyl flux toward TAG is canalized through acyl-CoA-independent routes, but the relative contribution of the acyl-CoA-dependent and -independent routes is mostly unknown in castor bean (Brown et al. 2012).

An enzymatic activity that transferred an acyl group from a DAG molecule to another, generating TAG and monoacylglycerol (MAG) has been reported in vitro (Stobart et al. 1997). This acyl-CoA-independent mechanism for TAG synthesis was attributed to the so-called diacylglycerol:diacylglycerol transacylase (DGTA). Although this activity has also been detected in vitro in other plants, neither the gene nor the enzyme responsible have been isolated, and it could be due to residual transacylation by PDAT using DAG instead of PC as the acyl donor.

11.3.6 Exclusion of Ricinoleic Acid from Membrane Lipids

It is well-known that RA, the main component of castor oil, as well as other unusual fatty acids such as epoxylated fatty acids found in certain plant species, are incompatible with cell membrane integrity. TAG is synthesized in different castor bean tissues besides the seed, yet those tissues' fatty acid profiles are very different those observed in seed. While seed TAGs are mostly composed of RA, this fatty acid is completely absent in other tissues (Brown et al. 2012), where it is efficiently excluded from PC, the molecule where RA is synthesized.

The main reason why RA is not found in other plant organs is because FAH12 is not expressed in any tissue except the seed endosperm and cotyledon (Brown et al. 2011, 2012). However, particular mechanisms are required in seed tissues to ensure RA exclusion from the synthesis of membrane lipids.

For those plants that accumulate unusual fatty acids in their oils, it was initially hypothesized that CPT prevented the incorporation of unusual fatty acids into PC by discriminating against DAG species that contained such fatty acids (Bafor et al. 1990). Under this hypothesis, CPT would play a role in excluding DAG molecular species with membrane-incompatible fatty acids from being used for PC synthesis while DGAT would have a preference for those same precursors (Vogel and Browse 1996). However, no CPT activity with any substantial ability to discriminate against DAG molecular species containing unusual fatty acids could be detected in four oilseed plants (safflower, castor bean, rapeseed, and *Cuphea lanceolata*) (Vogel and Browse 1996).

Furthermore, heterologous *RcFAH12* expression in some plants showed that although RA was ubiquitously synthesized, and it was always excluded from membrane lipids (van de Loo et al. 1995; Broun and Somerville 1997; McKeon et al. 1999; Chen et al. 2016b). This indicated the existence of a general mechanism for excluding unusual fatty acids from membrane lipids, probably because such fatty acids disturbed the membrane lipid bilayer. This mechanism is at least in part, carried out by LPA2, which efficiently removes "rare" FAs from PC (McKeon et al. 1999). It has also been shown that LPCAT can catalyze the reverse reaction (from PC to LPC plus acyl-CoA). Different LPCATs from different plant species including castor bean showed clear preference for RA release from PC (producing Ric-CoA plus LPC—the reverse reaction) while showing very low activity with Ric-CoA (producing PC esterified with RA) for the forward reaction (Bates 2016). As castor bean cannot accommodate RA-containing phospholipids in its plasma membranes, the preference showed by LPCAT would have a double effect: on one hand, it would favor the release of RA

from PC (a membrane lipid component) and on the other hand, it would consistently avoid RA incorporation into PC. In species that do not normally produce unusual fatty acids, the high reverse activity of LPCAT with HFA would represent a more general evolutionary adaptation to remove damaged (e.g., oxidized) fatty acids from membrane lipids because they may have an adverse effect on the membrane structure and function (Bates and Browse 2012; Bates 2016).

Following RA or "rare" FA release, it is either incorporated into TAG or catabolized through the β-oxidation process, both of which pathways occur very rapidly (Brown and Somerville 1997; Bates and Browse 2011; McKeon 2016).

In summary, efficient RA exclusion from membrane lipids in castor bean seems to be the result of a general mechanism that evolved in plants to warrant the exclusion of "rare" or damaged fatty acids from the membrane lipid bilayer. This mechanism would be accomplished using two main metabolic effectors: PLA2 and LPCAT. The excluded FAs would be efficiently incorporated into TAG, or quickly eliminated by β-oxidation.

11.3.7 TAG Compartmentalization and Membrane Synthesis Processes

Some years ago, Vogel and Browse (1996) suggested that two separate DAG pools existed: one devoted to TAG synthesis and the other used to generate membrane lipids. More recently, metabolic flux experiments performed in soybean (Bates et al. 2009) indicated that de novo and PC-derived DAG pools are separated. This evidence, in addition to the previously discussed mechanism to exclude unusual fatty acids from membrane lipids, strongly supports the idea that two separate lipid substrate pools exist, either for TAG synthesis or to produce membrane components. The way in which this lipid compartmentalization is achieved is unknown, but it might involve some form of separation of the lipid biosynthesis machinery in different "micro-domains" of the endoplasmic reticulum (ER) membrane (Shockey et al. 2006), besides unusual FA-specific enzymes described in previous sections. It is plausible that such micro-domains would be devoid of FAH12 and would contain membrane-editing enzymes (e.g., FAD2, FAD3, etc.).

Results from a recent study that analyzed a high-oleic acid mutant of castor bean (Venegas-Calerón et al. 2016) could be interpreted to provide further support for the micro-domain hypothesis. The reported high-oleic acid phenotype was due to a loss-of-function mutation in the *RcFAH12* gene, which led to a substantial reduction of RA production in mature seeds, as well as an anomalously high OA content. However, other fatty acids derived from OA desaturation (e.g., 18:2 and 18:3) were not augmented in the mutant seed (Venegas-Calerón et al. 2016). Although these results could be compatible with a common DAG pool for both TAG and membrane lipids production, the accumulated OA should produce a concomitant increase in OA-desaturated derivatives due to desaturase reactions, which was not observed. Another interpretation of these results could be that impairment of FAH12 affected only a discrete micro-domain where TAG is synthesized and reduced desaturase activity might be expected, resulting in minimal alteration of the amount of OA-desaturated derivatives in TAGs.

11.4 Castor Bean Biotechnology

The economic importance of castor oil is currently boosting research in diverse areas using different approaches. One of the main biotechnological approaches for castor oil production is the transfer of the necessary metabolic machinery to common oilseed crops in order to convert established oleaginous crops into castor oil producers. This would solve the problems posed by castor bean such as the presence of ricin and allergens, poor agronomic performance, and reduced adaptation to temperate climates. Some important achievements have already been made in this regard, while other biotechnological approaches are also being considered. For instance, castor bean improvement through genetic engineering to make it a friendly crop could be achieved by suppressing the

synthesis of ricin and allergens, or by enhancing desirable agronomic traits. Another possibility that merits investigation is the modification of the fatty acids profile looking for other specialty oils with unusual fatty acids in the castor bean.

11.4.1 Transfer of Castor Oil Machinery to Oil Crops

The first attempts at transgenic production of RA were performed in model plants such as Arabidopsis and *Nicotiana* as a proof of concept, to show that this strategy might work. A report on *RcFAH12* cloning also included testing of its heterologous expression in *Nicotiana tabacum* controlled by the cauliflower mosaic virus 35S promoter. The results were rather disappointing since RA was produced in the seed at a very low level (0.1%) and was not detectable in any other organ (van de Loo et al. 1995). Similar experiments in Arabidopsis showed the production HFA (mostly RA) reaching only 4% of the total fatty acids (TFAs) (Broun and Somerville 1997).

The low levels of RA production in transgenic plants were attributed, at least in part, to the poor performance of the 35S promoter, which was not seed-specific as the endogenous *RcFAH12* promoter. Heterologous expression of *RcFAH12* under the control of the seed-specific napin promoter improved RA production in transgenic Arabidopsis, although the highest accumulation of hydroxy fatty acids (including 18, 20, and 22 carbon FA) reached only about 17% of total seed lipids (Broun and Somerville 1997). Similar experiments in canola achieved a maximum of 20% of hydroxy fatty acids (McKeon 2016). It became apparent from these results that there was a limit to the amount of compositional change that might be brought about by expression of single heterologous gene (McKeon et al. 1999).

Further metabolic engineering of Arabidopsis, with both castor bean *RcFAH12* and *RcDGAT2* genes, resulted in a significant increase of RA in the seed oil, which reached 27% of TFA (Burgal et al. 2008). Nevertheless, this was far less than the amount accumulated in castor bean seed and therefore, commercially inviable.

Expression of castor bean oil biosynthesis genes has also been attempted in *Lesquerella* (*Physaria fendleri*, formerly *Lesquerella fendleri* (Gray) Wats.) by introduction of the *RcLPAT2* gene. In this Brassicaceae species, the major FA is lesquerolic acid (55–60% of the total seed fatty acids), another hydroxylated FA (20:1OH). As expected, several transgenic lines showed a significant RA increase in the *sn*-2 position of TAG (Chen et al. 2016a). However, although there was an increase in RA at the *sn*-2 position of TAG, the maximum level of RA attained in lines homozygous for the transgene was as low as 5%. This was probably due to the inability of endogenous acyltransferases to incorporate RA to other positions on the glycerol backbone.

The restrictions observed in transgenic Arabidopsis in accumulating triricinolein contrast with an accumulation of about 70% for this molecular species in castor oil. The information available shows that the limited ability of RA production in transgenic plants does not depend only on appropriate *RcFAH12* expression but also on a suite of other genes that are required to emulate the natural capability of castor bean's metabolism (Burgal et al. 2008; McKeon 2016).

The reasons for this limitation are differences in the TAG biosynthetic pathways between Arabidopsis, which uses PC-derived DAG (Lu et al. 2009; Bates and Browse 2011), and castor bean, which mainly uses de novo DAG (Bafor et al. 1991). It has been demonstrated that the production of diricinoleoyl diacylglycerol (2RA-DAG) through the predominantly PC-derived DAG pathway of TAG synthesis in Arabidopsis is a major limiting factor (Bates and Browse 2011).

The above example shows that the relative fluxes in TAG synthesis from de novo DAG or PC-derived DAG are important and should be taken into consideration for oilseed genetic engineering. The problem can also be extrapolated to other oilseed crops such as soybean, where it was shown that >95% of de novo DAG also fluxes through PC prior to TAG synthesis (Bates et al. 2009). Therefore, DAG flux through PC for TAG synthesis may represent a bottleneck for engineering many oilseed plants with different kinds of unusual FAs.

Future engineering strategies that focus on replacing endogenous de novo DAG synthesis enzymes with castor bean RA-specific homologs may result greater quantities of de novo 2RA-DAG in transgenic oilseeds. However, the flux of de novo DAG through PC may need to be circumvented in transgenic plants utilizing a PC-derived DAG pathway to reduce the HFA-containing de novo DAG turnover and make more de novo DAG available for direct conversion to TAG. An interesting alternative would be to use oilseed species in which TAG synthesis proceeds mainly from the de novo DAG reservoir, similar to that in castor bean (Bates and Browse 2012).

Other efforts have been directed toward the introduction of the castor oil biosynthesis machinery into industrial microorganisms using different genes and approaches.

The fungus *Claviceps purpurea* possesses a bifunctional enzyme, able to act as a delta-12 desaturase as well as a hydroxylase. It is encoded by a single gene called *CpFAH12*. As a hydroxylase, the CpFAH12 enzyme can synthetize RA and, therefore, its gene has been expressed in other species in order to produce RA. Hence, the oleaginous filamentous fungi *Ashbya gossypii* was transformed with the *CpFAH12* gene, although minimal RA production was obtained while there was a significant accumulation of linoleic acid LA (18:2). This showed that desaturase activity generating LA from OA dominated the bifunctional nature of CpFAH12 (Bardone et al. 2016), at least under the conditions used for fungus cultivation.

In another attempt at producing RA in an industrial yeast, *Pichia pastoris* was simultaneously engineered with both the *CpFAH12* and the *CpDGAT1* genes from *C. purpurea*. Co-expression of these genes led to RA accumulation of >50% of the total FAs, but mostly in the free FA form (Meesapyodsuk et al. 2015). This was surprising given that it is usually assumed that free FAs are toxic for cells.

A similar approach was also carried out in the diatom *Chaetoceros gracilis*, which was transformed with the *CpFAH12* gene. This oleaginous microalga showed an RA accumulation of up to 2.2 pg/cell (Kajikawa et al. 2016). However, RA was present predominantly as monoestolide triacylglycerol, in which one RA molecule was esterified to the α position of the glycerol backbone and it was further esterified at its hydroxyl group with a fatty acid or a second RA moiety, or 1-OH TAG, where RA was esterified to the glycerol backbone.

11.4.2 Biotechnological Improvement of Castor Bean

The main focus of the efforts toward the development of an improved castor bean cultivar is to solve the problems that prevent the use of this crop at a larger scale. Elimination of the ricin toxin and the allergens produced in its seeds could make castor bean a suitable oleaginous plant. Its use would even be conceivable as a biotechnological platform for the synthesis of valuable products without interfering with food crops and avoiding concerns regarding flow of transgenes to food crops.

The genes involved in the production of ricin and the main protein allergens are known, and they have been annotated on the publicly available castor bean genome sequence. Moreover, genetic engineering strategies that have been established in other plants could be used to silence the expression or knock out undesirable castor bean genes such as ricin and allergens.

One of the main obstacles in achieving this, however, is the availability of an efficient transformation methodology for castor bean. Even though stable nuclear transformation of castor bean has been reported (Sujatha et al. 2008), its efficiency is rather low, making it difficult to apply state-of-the-art gene editing technologies to knock out multigene families, such as that of ricin.

11.5 Concluding Remarks and Future Prospects

TAG accumulation enriched in RA requires the coordinated flux of OA into PC for hydroxylation, and the subsequent channeling of RA out of PC and into TAG. As we have seen, the castor oil metabolic machinery is particularly well suited to this end (McKeon 2016). Some of the identified steps leading to high ricinoleate production and incorporation into triacylglycerol include reactions catalyzed by LPCAT, FAH12, PLA2, LPAT, and DGAT, meaning that many of the enzymes which are involved in TAG synthesis have evolved in this species to favor oil enrichment in RA.

Even though knowledge of plant TAG biosynthesis has increased considerably over the last 20 years, there are still many gaps that make the control of TAG metabolic engineering difficult in plants (Bates 2016). Some of the most critical challenges include controlling the acyl flux in various fatty acid modifications and the DAG pool used to produce TAG. Variations in fluxes through the diverse pathways for TAG synthesis in different plant species add additional complexity when attempting to engineer FA composition in one plant to make it similar to another (Bates and Browse 2011).

Although knowledge of castor oil biosynthesis has greatly advanced, further research will facilitate the design of rational engineering of TAG synthesis in castor bean and other oleaginous crops. The availability of the castor bean genome sequence constitutes a critical tool for accomplishing this aim.

Increased understanding of the regulation of TAG biosynthesis could provide a qualitative shift in this research field. Further insights on the possible existence of different discrete micro-domains in the ER for storage and membrane lipids, along with a characterization of the metabolic fluxes between both compartments, would greatly advance our general understanding of seed lipid metabolism.

Progress on other areas is also necessary, such as the identification of genes encoding some of the key TAG biosynthetic enzymes that are still unknown in castor bean, such as PLA2 and LACS, which may play an important role in biosynthesizing castor oil (Lin et al. 1998; McKeon et al. 1999), but remain uncharacterized. In addition, an in-depth analysis of castor bean GPAT and its putative gene (*RcGPAT9*) will be important for understanding this key oil biosynthesis step, and its possible role in channeling RA into TAG.

Significant progress can undoubtedly be made in castor bean genetic transformation, simply by increasing the efficiency of available methodologies or by developing new strategies. A substantial knowledge of castor oil biosynthesis combined with optimized genetic engineering tools for castor bean will greatly advance this crop's biotechnological potential.

References

Arroyo-Caro JM, Chileh T, Alonso DL, García-Maroto F (2013a) Molecular characterization of a lysophosphatidylcholine acyltransferase gene belonging to the MBOAT family in *Ricinus communis* L. Lipids 48: 663–674. https://doi.org/10.1007/s11745-013-3797-z

Arroyo-Caro JM, Chileh T, Kazachkov M, Zou J, Alonso DL et al (2013b) The multigene family of lysophosphatidate acyltransferase (LPAT)-related enzymes in *Ricinus communis*. Cloning and molecular characterization of two *LPAT* genes that are expressed in castor seeds. Plant Sci 199–200:29–40. https://doi.org/10.1016/j.plantsci.2012.09.015

Arroyo-Caro JM, Mañas-Fernández A, Alonso DL, García-Maroto F (2016) Type I diacylglycerol acyltransferase (MtDGAT1) from *Macadamia tetraphylla*: cloning, characterization, and impact of its heterologous expression on triacylglycerol composition in yeast. J Agric Food Chem 64:277–285. https://doi.org/10.1021/acs.jafc.5b04805

Bafor M, Jonsson L, Stobart AK, Stymne S, Stobart K et al (1990) Regulation of triacylglycerol biosynthesis in embryos and microsomal preparations from the developing seeds of *Cuphea lanceolata*. Biochem J 272:31–38

Bafor M, Smith MA, Jonsson L, Stobart K, Stymne S (1991) Ricinoleic acid biosynthesis and triacylglycerol assembly in microsomal preparations from developing castor-bean (*Ricinus communis*) endosperm. Biochem J 280:507–514

Bardone E, Bravi M, Keshavarz T, Lozano-Martínez P, Ledesma-Amaro R et al (2016) Engineering *Ashbya gossypii* for the production of ricinoleic and linoleic

acid. Chem Eng Trans 49:253–258. https://doi.org/10.3303/CET1649043

Bates PD (2016) Understanding the control of acyl flux through the lipid metabolic network of plant oil biosynthesis. Biochim Biophys Acta 1861:1214–1225. https://doi.org/10.1016/j.bbalip.2016.03.021

Bates PD, Browse JA (2011) The pathway of triacylglycerol synthesis through phosphatidylcholine in *Arabidopsis* produces a bottleneck for the accumulation of unusual fatty acids in transgenic seeds. Plant J 68:387–399. https://doi.org/10.1111/j.1365-313X.2011.04693.x

Bates PD, Browse JA (2012) The significance of different diacylgycerol synthesis pathways on plant oil composition and bioengineering. Front Plant Sci 3:1–11. https://doi.org/10.3389/fpls.2012.00147

Bates PD, Durrett TP, Ohlrogge JB, Pollard M (2009) Analysis of acyl fluxes through multiple pathways of triacylglycerol synthesis in developing soybean embryos. Plant Physiol 150:55–72. https://doi.org/10.1104/pp.109.137737

Bates PD, Fatihi A, Snapp AR, Carlsson AS, Browse J et al (2012) Acyl editing and headgroup exchange are the major mechanisms that direct polyunsaturated fatty acid flux into triacylglycerols. Plant Physiol 160:1530–1539. https://doi.org/10.1104/pp.112.204438

Broun P, Somerville CR (1997) Accumulation of ricinoleic, lesquerolic, and densipolic acids in seeds of transgenic *Arabidopsis* plants that express a fatty acyl hydroxylase cDNA from castor bean. Plant Physiol 113:933–942

Brown AP, Carnaby S, Brough CL, Brazier M, Slabas AR (2002) *Limnanthes douglasii* lysophosphatidic acid acyltransferases: immunological quantification, acyl selectivity and functional replacement of the *Escherichia coli plsC* gene. Biochem J 364:795–805. https://doi.org/10.1042/BJ20020071

Brown AP, Kroon JT, Topping JF, Robson JL, Simon WJ et al (2011) Components of complex lipid biosynthetic pathways in developing castor (*Ricinus communis*) seeds identified by MudPIT analysis of enriched endoplasmic reticulum. J Proteome Res 10:3565–3577. https://doi.org/10.1021/pr2002066

Brown AP, Kroon JT, Swarbreck D, Febrer M, Larson TR et al (2012) Tissue-specific whole transcriptome sequencing in castor, directed at understanding triacylglycerol lipid biosynthetic pathways. PLoS ONE 7:e30100. https://doi.org/10.1371/journal.pone.0030100

Burgal J, Shockey J, Lu C, Dyer J, Larson T et al (2008) Metabolic engineering of hydroxy fatty acid production in plants: RcDGAT2 drives dramatic increases in ricinoleate levels in seed oil. Plant Biotechnol J 6:819–831. https://doi.org/10.1111/j.1467-7652.2008.00361.x

Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J et al (2010) Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotechnol 28:951–956. https://doi.org/10.1038/nbt.1674

Chen GQ, Turner C, He X, Nguyen T, McKeon TA et al (2007) Expression profiles of genes involved in fatty acid and triacylglycerol synthesis in castor bean (*Ricinus communis* L.). Lipids 42:263–274. https://doi.org/10.1007/s11745-007-3022-z

Chen GQ, Johnson K, Morales E, Mackey B, Lin JT (2016a) Rapid development of a castor cultivar with increased oil content. Ind Crops Prod 94:586–588. https://doi.org/10.1016/j.indcrop.2016.09.020

Chen GQ, van Erp H, Martin-Moreno J, Johnson K, Morales E et al (2016b) Expression of castor LPAT2 enhances ricinoleic acid content at the sn-2 position of triacylglycerols in *Lesquerella* seed. Int J Mol Sci 17:1–14. https://doi.org/10.3390/ijms17040507

Dahlqvist A, Stahl U, Lenman M, Banas A, Lee M et al (2000) Phospholipid:diacylglycerol acyltransferase: an enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. Proc Natl Acad Sci USA 97:6487–6492

FAOSTAT (Food and Agriculture Organization of the United Nations). http://www.fao.org/faostat

Frentzen M (1998) Acyltransferases from basic science to modified seed oils. Fett-Lipid 100:161–166

Haslam RP, Sayanova O, Kim HJ, Cahoon EB, Napier JA (2016) Synthetic redesign of plant lipid metabolism. Plant J 87:76–86. https://doi.org/10.1111/tpj.13172

He X, Chen GQ, Lin JT, McKeon TA (2004a) Regulation of diacylglycerol acyltransferase in developing seeds of castor. Lipids 39:865–871

He X, Turner C, Chen GQ, Lin JT, McKeon TA (2004b) Cloning and characterization of a cDNA encoding diacylglycerol acyltransferase from castor bean. Lipids 39:311–318. https://doi.org/10.1007/s11745-004-1234-2

He X, Chen GQ, Lin JT, McKeon TA (2006) Diacylglycerol acyltransferase activity and triacylglycerol synthesis in germinating castor seed cotyledons. Lipids 41:281–285

He X, Chen GQ, Kang ST, McKeon TA (2007) *Ricinus communis* contains an acyl-CoA synthetase that preferentially activates ricinoleate to its CoA thioester. Lipids 42:931–938. https://doi.org/10.1007/s11745-007-3090-0

Hu Z, Ren Z, Lu C (2012) The phosphatidylcholine diacylglycerol cholinephosphotransferase is required for efficient hydroxy fatty acid accumulation in transgenic *Arabidopsis*. Plant Physiol 158:1944–1954. https://doi.org/10.1104/pp.111.192153

Kajikawa M, Abe T, Ifuku K, Furutani K, Yan D et al (2016) Production of ricinoleic acid-containing monoestolide triacylglycerides in an oleaginous diatom, *Chaetoceros gracilis*. Sci Rep 6:36809. https://doi.org/10.1038/srep36809

Kim HU, Li Y, Huang AHC (2005) Ubiquitous and endoplasmic reticulum-located lysophosphatidyl acyltransferase, LPAT2, is essential for female but not male gametophyte development in *Arabidopsis*. Plant Cell 17:1073–1089. https://doi.org/10.1105/tpc.104.030403

Knutzon DS, Lardizabal KD, Nelsen JS, Bleibaum JL, Davies HM et al (1995) Cloning of a coconut endosperm cDNA encoding a 1-acyl-sn-glycerol-3-phosphate

acyltransferase that accepts medium-chain-length substrates. Plant Physiol 109:999–1006

Kroon JT, Wei W, Simon WJ, Slabas AR (2006) Identification and functional expression of a type 2 acyl-CoA:diacylglycerol acyltransferase (DGAT2) in developing castor bean seeds which has high homology to the major triglyceride biosynthetic enzyme of fungi and animals. Phytochemistry 67:2541–2549. https://doi.org/10.1016/j.phytochem.2006.09.020

Lager I, Yilmaz JL, Zhou X-R, Jasieniecka K, Kazachkov M et al (2013) Plant acyl-CoA:lysophosphatidylcholine acyltransferases (LPCATs) have different specificities in their forward and reverse reactions. J Biol Chem 288:36902–36914

Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V et al (2013) Acyl-lipid metabolism. In: The Arabidopsis book. p e0161

Lin JT, Woodruff CL, Lagouche OJ, McKeon TA, Stafford AE et al (1998) Biosynthesis of triacylglycerols containing ricinoleate in castor microsomes using 1-acyl-2-oleoyl-sn-glycero-3-phosphocholine as the substrate of oleoyl-12-hydroxylase. Lipids 33:59–69

Lu C, Wallis JG, Browse JA (2007) An analysis of expressed sequence tags of developing castor endosperm using a full-length cDNA library. BMC Plant Biol 7:42. https://doi.org/10.1186/1471-2229-7-42

Lu C, Xin Z, Ren Z, Miquel MF, Browse JA (2009) An enzyme regulating triacylglycerol composition is encoded by the *ROD1* gene of *Arabidopsis*. Proc Natl Acad Sci USA 106:18837–18842. https://doi.org/10.1073/pnas.0908848106

Mañas-Fernández A, Vilches-Ferrón MA, Garrido-Cárdenas JA, Belarbi EH, Alonso DL et al (2009) Cloning and molecular characterization of the acyl-CoA:diacylglycerol acyltransferase 1 (DGAT1) gene from *Echium*. Lipids 44:555–568. https://doi.org/10.1007/s11745-009-3303-9

McKeon TA (2016) Castor (*Ricinus communis* L.). In: Industrial oil crops. AOCS Press. Published by Elsevier Inc., pp 75–112

McKeon TA, He X (2015) Castor diacylglycerol acyltransferase type 1 (DGAT1) displays greater activity with diricinolein than *Arabidopsis* DGAT1. Biocatal Agric Biotechnol 4:276–278. https://doi.org/10.1016/j.bcab.2015.01.005

McKeon T, Lin J, Goodrich-Tanrikulu M, Stafford A (1997) Ricinoleate biosynthesis in castor microsomes. Ind Crops Prod 6:383–389. https://doi.org/10.1016/S0926-6690(97)00029-0

McKeon TA, Lin JT, Stafford AE (1999) Biosynthesis of ricinoleate in castor oil. Adv Exp Med Biol 464:37–47

Meesapyodsuk D, Chen Y, Ng SH, Chen J, Qiu X (2015) Metabolic engineering of *Pichia pastoris* to produce ricinoleic acid, a hydroxy fatty acid of industrial importance. J Lipid Res 56:2102–2109. https://doi.org/10.1194/jlr.M060954

Millar AA, Smith MA, Kunst L (2000) All fatty acids are not equal: discrimination in plant membrane lipids. Trends Plant Sci 5:95–101

Moreau RA, Stumpf PK (1981) Recent studies of the enzymic synthesis of ricinoleic acid by developing castor beans. Plant Physiol 67:672–676

Shockey JM, Gidda SK, Chapital DC, Kuan J-C, Dhanoa PK, Bland JM, Rothstein SJ, Mullen RT, Dyer JM (2006) Tung tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum. Plant Cell 18:2294–2313. https://doi.org/10.1105/tpc.106.043695

Shockey J, Regmi A, Cotton K, Adhikari N, Browse J et al (2016) Identification of *Arabidopsis GPAT9* (At5g60620) as an essential gene involved in triacylglycerol biosynthesis. Plant Physiol 170:163–179. https://doi.org/10.1104/pp.15.00025

Singer SD, Chen G, Mietkiewska E, Tomasi P, Jayawardhane K et al (2016) *Arabidopsis* GPAT9 contributes to synthesis of intracellular glycerolipids but not surface lipids. J Exp Bot 67:4627–4638. https://doi.org/10.1093/jxb/erw242

Snyder CL, Yurchenko OP, Siloto RMP, Chen X, Liu Q et al (2009) Acyltransferase action in the modification of seed oil biosynthesis. Nat Biotechnol 26:11–16

Stobart K, Mancha M, Lenman M, Dahlqvist A, Stymne S et al (1997) Triacylglycerols are synthesised and utilized by transacylation reactions in microsomal preparations of developing safflower (*Carthamus tinctorius* L.) seeds. Planta 203:58–66

Stymne S, Stobart K (1984) Evidence for the reversibility of the acyl-CoA:lysophosphatidylcholine acyltransferase in microsomal preparations from developing safflower (*Carthamus tinctorius* L.) cotyledons and rat liver. Biochem J 223:305–314

Sujatha M, Reddy TP, Mahasi MJ (2008) Role of biotechnological interventions in the improvement of castor (*Ricinus communis* L.) and *Jatropha curcas* L. Biotechnol Adv 26:424–435. https://doi.org/10.1016/j.biotechadv.2008.05.004

Tomosugi M, Ichihara K, Saito K (2006) Polyamines are essential for the synthesis of 2-ricinoleoyl phosphatidic acid in developing seeds of castor. Planta 223:349–358. https://doi.org/10.1007/s00425-005-0083-1

van de Loo FJ, Broun P, Turner S, Somerville CR (1995) An oleate 12-hydroxylase from *Ricinus communis* L. is a fatty acyl desaturase homolog. Proc Natl Acad Sci USA 92:6743–6747. https://doi.org/10.1073/pnas.92.15.6743

van Erp H, Bates PD, Burgal J, Shockey JM, Browse JA et al (2011) Castor phospholipid: diacylglycerol acyltransferase facilitates efficient metabolism of hydroxy fatty acids in transgenic arabidopsis. Plant Physiol 155:683–693

Venegas-Calerón M, Sánchez R, Salas JJ, Garcés R, Martínez-Force E (2016) Molecular and biochemical characterization of the OLE-1 high-oleic castor seed (*Ricinus communis* L.) mutant. Planta 244:245–258. https://doi.org/10.1007/s00425-016-2508-4

Vinay R, Viswanath LCK, Maples R, J BJ (2016) Castor oil: properties, uses, and optimization of processing

parameters in commercial production. Lipid Insights 9:1–12. https://doi.org/10.4137/lpi.s40233

Vogel G, Browse JA (1996) Cholinephosphotransferase and diacylglycerol acyltransferase (substrate specificities at a key branch point in seed lipid metabolism). Plant Physiol 110:923–931

Yadav P, Anjani K (2017) Assessment of variation in castor genetic resources for oil characteristics. J Am Oil Chem Soc 94:611–617. https://doi.org/10.1007/s11746-017-2961-7

Yu K, McCracken CT Jr, Li R, Hildebrand DF (2006) Diacylglycerol acyltransferases from *Vernonia* and *Stokesia* prefer substrates with vernolic acid. Lipids 41:557–566

12 Ricin and RCA—The Enemies Within Castor (*Ricinus communis* L.): A Perspective on Their Biogenesis, Mechanism of Action, Detection Methods and Detoxification Strategies

Mohd. Ashraf Ashfaq, P. Soma Sekhar Reddy, Ch. Anil Kumar, Velu Mani Selvaraj and V. Dinesh Kumar

Abstract

The main impediment in utilizing the de-oiled meal of castor bean (*Ricinus communis* L.) as animal feed is the presence of two toxic proteins, ricin (also known as RCA II and RCA_{60}) and RCA (*R. communis* agglutinin, otherwise known as RCA I and RCA_{120}) found in the endosperm of mature seeds. Their presence compromises with the high protein content (around 25% with balanced amino acid composition) in the leftover oilcake which otherwise could find use as an excellent animal feed. This review is an attempt at providing insights into the biology and biogenesis of ricin and RCA and various approaches for reducing or eliminating ricin and RCA in this oil crop.

Mohd.Ashraf Ashfaq · P. Soma Sekhar Reddy · Ch. Anil Kumar · V. M. Selvaraj · V. Dinesh Kumar (✉)
ICAR-Indian Institute of Oilseeds Research, 500030 Rajendranagar, Hyderabad, India
e-mail: vdinesh.kumar1@icar.gov.in

Mohd.Ashraf Ashfaq
ICAR-Indian Institute of Rice Research, 500030 Rajendranagar, Hyderabad, India

P. Soma Sekhar Reddy · Ch.Anil Kumar
Department of Genetics, Osmania University, 500007 Hyderabad, India

Mohd.Ashraf Ashfaq · V. M. Selvaraj
Department of Plant Sciences, University of Hyderabad, 500046 Hyderabad, India

12.1 Introduction

Castor bean (*R. communis* L.) is a commercially important non-edible oilseed crop in the dicotyledonous angiosperm family Euphorbiaceae. It is an important source of vegetable oil and medicinal oil. Dehydrated castor oil is used in the paint and varnish industry, manufacture of a wide range of sophisticated products like nylon fibres, jet engine lubricants, hydraulic fluids, plastics, artificial leather, fibre optics, bulletproof glass, bone prostheses, and antifreeze for fuels and lubricants used in aircraft and space rockets (Scarpa and Guerci 1982; Ogunniyi 2006). Castor oil is unique among vegetable oils because of the presence of ricinoleic acid, a hydroxy fatty acid that constitutes up to 90% of the oil. The presence of long (18 C) fatty aliphatic carbon chain, a carboxyl group, a hydroxyl group, and a double bond, as well as the proximity of the latter two offers exciting possibilities for many chemical reactions and for the production of a variety of oleo chemicals. Global demand for castor oil is 1 billion pounds annually, with an economic value of 500 million USD. Although the USA consumes more than 10% of the world's castor bean production, virtually none has been produced in the USA since the early 1970s. India, China and Brazil account for more than 90% of the world's castor bean supply. India is yet to levy the potential of its monopoly as mostly raw oil and a few value-added products are exported.

C. Kole and P. Rabinowicz (eds.), *The Castor Bean Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97280-0_12

Castor cake (or oilcake) is a by-product of the milling industry, which accounts for 60% of the crushed seed that can account for 0.669 million tons annually (considering an annual production of 1.115 million tons of castor bean seed as reported in 2008–2009). It is rich in protein (25%), sugar (25%) and minerals (20%). Furthermore, castor cake is a good source of vitamins like thiamine, beta-carotene and tocopherols. Thus, castor oilcake could be an excellent animal feed. Castor cake is useful as organic manure (N—6.0%, P_2O_5—2.5%, K_2O—2.5%). However, the presence of many toxic constituents renders castor cake non-edible making it less useful. The most toxic constituents of castor cake are ricin and RCA, whose concentrations can reach up to 125.5 mg/l00 g of oilcake. The cake is also rich in allergens, which, along with the presence of ricin, has resulted in the banning of castor bean cultivation in the USA and some European countries. The total elimination or inactivation of ricin, RCA and allergens from castor cake is extremely important before it can be considered for animal feed, fertilizer or wastewater pretreatment (as solid enzymatic preparation). Furthermore, if the residue does not possess any utility after the detoxification treatment, it may be discarded in a landfill without the added danger of soil and water contamination caused by the presence of ricin. The toxic constituents of castor bean provide nematicidal and termiticidal properties due to which castor cake is used as organic manure in plantation crops.

Plants are replete with thousands of proteins and small molecules, many of which are species specific, poisonous or dangerous. Over time, humans have learned to avoid dangerous plants or inactivate many toxic components in food plants. Down-regulating the genes responsible for toxic substances, which would result in lines with reduced levels of toxins, has been an objective in several breeding programmes. Inactivation at the genetic rather than physical or chemical level has many advantages, and classical genetic approaches have resulted in significant reduction of toxic content in plants. Traditionally, this has been achieved by careful selection of genotypes from the available genetic diversity. Alternatively, new variations could be generated through the use of mutagens. However, both these processes are slow and costly, and most often unsuccessful.

The capacity offered by genetic engineering in inactivating specific genes has opened up the possibility of altering the plant content in a far more precise manner than previously available. Different levels of intervention (altering genes coding for toxins/allergens or for enzymes, transporters or regulators involved in their metabolism) are possible, and there are several tools for inactivating genes, both direct (using chemical and physical mutagens, insertion of transposons and other genetic elements) and indirect (antisense RNA, RNA interference, microRNA, eventually leading to gene silencing). Each level/strategy has specific advantages and disadvantages (speed, costs, selectivity, stability, reversibility, frequency of desired genotype and regulatory regime).

The use of post-transcriptional gene silencing (or PTGS, explained in more detail later in this chapter) to improve crop value by eliminating undesirable traits can potentially address many problems in a faster and more precise way than conventional breeding. Multiple genes sharing stretches of identical sequence can be effectively silenced with a single RNAi (RNA interference) construct. This feature is especially important in crops harbouring multiple copies of the target gene (as in the case of ricin and RCA in castor bean) and in polyploid crops like wheat (Fu et al. 2007), which have multiple homeologous copies of each gene. Therefore, RNAi is coming of age as a useful and flexible tool to study gene function in polyploid species, although more research is necessary to establish the additional factors that determine a construct's silencing efficiency.

In this chapter, we will provide an overview of chemistry, biogenesis, mechanism of action of ricin and RCA followed by the efforts made towards detoxification of these toxic proteins.

12.2 Chemistry of Ricin and RCA

Ricin (also known as RCA II or RCA_{60}) and *R. communis* agglutinin (otherwise known as RCA I, RCA_{120} or simply RCA) are the two

highly toxic, endosperm-specific proteins of castor bean. These are essentially glycoproteins and belong to the plant lectin family. Lectins are proteins that specifically bind carbohydrates and frequently agglutinate cells in vitro. Ricin is a potent cytotoxin but a weak haemagglutinin, whereas RCA is a weak cytotoxin and a powerful haemagglutinin.

Ricin, one of the most toxic compounds produced in nature, belongs to the type II ribosome-inactivating protein (RIP) family, members of which are found in most plant genera as well as in some fungi and bacteria. It is a heterodimeric polypeptide consisting of a 32 KDa ricin toxin A (RTA)-chain—a potent inhibitor of 80 S ribosomes—and a 34 KDa carbohydrate-binding ricin toxin B (RTB)-chain, linked by a single disulphide bond.

Ricin is a cytosolic rather than a membrane-bound protein, as it shows no distinct pattern of hydrophobic and hydrophilic regions. RTA is a globular protein with a prominent and conspicuous binding site cleft; RTB is an elongated dumbbell-shaped protein with galactose-binding sites at both ends. X-ray crystallography has been employed to work out the three-dimensional structure of ricin and is refined and resolved at 2.5 Å (Montfort et al. 1987; Rutenber et al. 1991). This model detailed an accurate description of both the A-chain and the B-chain. Also, the structure of recombinant RTA has been resolved to 2.3 Å. Ricin A-chain is composed of 267 amino acid residues and has 28SrRNA N-glycosidase enzymatic activity. The active A-chain is approximately 30% helical and contains 7 alpha helices. It also contains about 15% beta structure, which consists of a free stranded beta sheet (Montfort et al. 1987). The B-chain is composed of 262 amino acid residues and is a lectin with affinity for galactoside binding (Katzin et al. 1991). RTB possesses two galactose-binding sites that are attracted to galactose-containing glycoproteins on the cell surface (Wiley and Oeltman 1991). The A- and B-chains are linked by a disulphide bond located at residue 259 of the A-chain and residue 4 of the B-chain (Montfort et al. 1987).

Being a glycoprotein, ricin possesses mannose-rich N-linked oligosaccharides as carbohydrate side chains. These functional side chains help in binding of the toxin to certain cell types with mannose receptors. In particular, ricin binds to mannose receptors of the cells of reticuloendothelial system (Wiley and Oeltman 1991). Ricin has sites with potential for binding to high mannose carbohydrate chains at asparagines 10 and 236 of the A-chain and asparagines 95 and 135 of the B-chain (Rutenber and Robertus 1991).

About eight conserved amino acids are of common occurrence in the family of RIPs. It is interesting to note that these conserved amino acids form the active site. These include Tyr 80, Tyr 123, Glu 177, Arg 180 and Trp 211. With respect to investigations into substrate binding, it has come to the fore that the substrate purine ring is stacked between the rings of the two conserved tyrosines. Arg 180 bonds to N3 of the ring, and Glu 177 stays near the ribose. Trp 211 makes no specific contacts with the adenine but may be important for the active site conformation or may interact with a larger polynucleotide substrate.

RCA is a four-chained polypeptide comprising of two A-chains and two B-chains, slightly less toxic than ricin and causes agglutination of red blood cells (RBCs) in mammals. RCA I is a tetramer composed of two ricin-like heterodimers held together by non-covalent forces. The A-chain of RCA has a 93% similarity to RTA, while the two B-chains are 84% similar to RTB (Roberts et al. 1985). While RTB has two glycans, RCA B-chain possesses three, with the additional glycan containing a fucose residue (Lord 1985a, b). This additional oligosaccharide accounts for the higher apparent molecular mass of RCA B-chain (~36 kDa) compared with RTB (34 kDa). Two of the 18 differences between the respective A-chains involve the substitution of cysteine residues in RCA, one of which (Cys 156) is reported to form a disulphide bond with an adjacent molecule to generate a mature ~128 kDa lectin with the subunit arrangement B–A–A–B (Cawley and Houston 1979; Sweeney et al. 1997). In contrast to ricin,

RCA interacts solely with galactose and not with N-acetylgalactosamine, a feature that permits convenient separation of these closely related lectins using affinity columns for protein purification (Nicolson et al. 1974).

12.3 Localization of Ricin and RCA

Ricin is produced in a tissue-specific manner in the endosperm during the post-testa maturation stage (Roberts and Lord 1981). Ricin, its isoforms, storage albumins, and crystalloid proteins are usually targeted for storage in an organelle called the protein body in the mature seed (Tulley and Beevers 1976; Youle and Huang 1976). The protein body is analogous to any other vacuolar compartment in the cell. Protein bodies are also present in the cotyledons, where ricin and RCA are part of the albumin fraction. Cotyledons contain much less lectins than the endosperm. Within the protein bodies, ricin accumulates to around 5% of the total particulate protein and is usually degraded within a few days after seed germination (Frigerio and Roberts 1998).

A study by Tregear and Roberts (1992) examined ricin and RCA mRNA levels during castor bean seed development using Northern blot analysis, but it could not discriminate between ricin and RCA transcripts due to their high degree of sequence identity. The developmental expression of ricin and RCA was also studied in seed using RT-PCR and Northern blot analyses (Chen et al. 2005). Using seed coat colour and endosperm volume, they divided the seed development into different phases; endosperm tissue starts developing at about 26 days after pollination (DAP), continues to develop until 54 DAP and by 61 DAP, seeds mature and desiccate. Analyses of ricin and RCA genes at transcript level have indicated that their expression coincides with the developmental phases of the endosperm with mRNA signal detected after 12 DAP, peaking between 26 and 54 DAP, the period when endosperm expands and occupies the entire seed, and then starts dropping with only traces of mRNA signals present by the time seed matures at 61 DAP. The RT-PCR assay/restriction fragment analysis reported by Chen et al. (2005) is definitive and a highly sensitive method to detect the expression of ricin or RCA genes at the transcriptional level. Thus, this technique provides a method to screen the effectiveness of different ricin and RCA gene silencing constructs in reducing the transcript levels of ricin and RCA in developing castor bean seeds in individual transformants.

Barnes et al. (2009) have shown that ricin is present in significant amounts in castor bean seed after 28 days post-pollination, and it disappears from the plant approximately 6 days after radicle emergence. Because of its late synthesis in the seed and early disappearance from the seedling, ricin likely does not play a defence role in the mature plant but rather ricin may only provide protection from predators to the mature seed.

12.4 Mechanism of Action of Ricin and RCA

As other RIP proteins, ricin is a toxic N-glycosidase that depurinates large 28SrRNA molecules and renders the ribosomes incapable of sustaining further translation. Endo and Tsurugi (1987) showed that the enzymatic activity of ricin was an N-glycosidation that removes a specific adenine corresponding to residue A 4324 in rat 28SrRNA. This adenine lies within a 14-nucleotide region known as a-sarcin loop and is conserved in large rRNAs from bacteria to humans. Irreversible modification of the target A residue blocks elongation factor (EF-1 and EF-2)-dependent GTPase activities and renders the ribosome unable to bind EF-2, thereby blocking translation. Thus, RTA is a 28SrRNA N-glycosidase (Endo and Tsurugi 1987), inactivating the ribosomes by depurinating adenine in a hairpin loop containing the tetranucleotide loop GAGA of 28SrRNA (Lord et al. 1994). A single A-chain molecule is capable of inactivating every ribosome in the cell thus, halting protein synthesis and culminating in cell death (Wiley and Oeltman 1991). Ricin has a Michaelis constant (K_m) of 0.1 mol/L for ribosomes and an

enzymatic constant (K_{cat}) of 1500/min. It is likely that the susceptible adenine base binds between tyrosine residues 80 and 123 while forming specific hydrogen bonds with the backbone carboxyl group and amino nitrogen atom of Val 81 and with the carboxyl group of Gly 12. In the hydrolysis, the leaving adenine is at least partially protonated by Arg 180, and Gln 177 may stabilize a putative oxycarbonium transition state, or more likely, act as a base to polarize the attacking water. However, it is thought that the substrate (ribosome) conformation is an important factor in its recognition by ricin (Lord et al. 1994). Therefore, ricin is not a nucleotide sequence-specific protein (Wiley and Oeltman 1991). It has also been shown that the sensitivity of ribosomes from different sources to the action of RTA (the A-chain of ricin protein that actually has the enzymatic activity for deadenylation) varies significantly. Thus, while mammalian ribosomes are most sensitive to the action of RTA, yeast ribosomes are slightly less sensitive while prokaryotic ribosomes are resistant as they do not contain 28SrRNA in their ribosomes (Lord et al. 1994).

12.5 Biogenesis of Ricin

The process of ricin biogenesis can be divided into three main phases: (i) co-translational import into and modification of ricin in the endoplasmic reticulum (ER), (ii) export into the Golgi complex and further modifications therein, and (iii) export into the protein bodies, maturation and storage. The process is illustrated in Fig. 12.1.

(i) **Co-translational import into and modification of ricin in the endoplasmic reticulum:** Preproricin is co-translationally translocated into the ER. The primary translation product is a 576 residue precursor containing an N-terminal 26-amino acid signal peptide for ER import followed by a 9-residue propeptide that influences import and glycosylation efficiency and is later cleaved (Ferrini et al. 1995; Jolliffe et al. 2006). Next to this propeptide is the RTA (267 residues) sequence separated from RTB (262 residues) by a 12 amino acid residue linker (Harley and Lord 1985; Lamb et al. 1985). Preproricin is co-translationally N-glycosylated (Lord 1985a), and multiple disulphide bonds are formed within RTB with an additional disulphide between the RTA and RTB sequences (Lord 1985b). It has been shown that the formation of disulphide bonds in ricin A-chain decreases the cytotoxicity of the ricin holotoxin (Argent et al. 1994). The signal peptide is cleaved upon entry into the ER to form proricin, which is catalytically inactive. This is because the sugar-binding RTB sterically hinders the catalytic site of RTA (Richardson et al. 1989). Indeed, this masking persists in the mature heterodimer (Montfort et al. 1987) but, during the entry into mammalian cells, the active site of RTA becomes exposed once RTA and RTB are reductively cleaved—an event that appears to occur in the ER of intoxicated cells (Spooner et al. 2004).

The fact that ricin molecules are synthesized as inactive precursors may explain how castor bean endosperm cells are able to produce, transport and store such large amounts of ricin even though endogenous ribosomes are susceptible to its action, albeit much less compared to animal ribosomes (Cawley et al. 1977; Harley and Beevers 1982; Taylor et al. 1994). In this way, should any precursor be inefficiently or incompletely imported, the hindered active site would prevent ribosome inactivation.

(ii). **Export into the Golgi apparatus and further modifications:** Radiolabeling experiments showed that proricin travels through the Golgi complex where the glycans on both chains become processed in a Golgi-specific manner (Lord 1985a). Indeed, one of the two RTA glycans (but neither of the two RTB glycans) contains an α-1,3 fucose linked to the proximal N-acetylglucosamine residue (Lord 1985b)

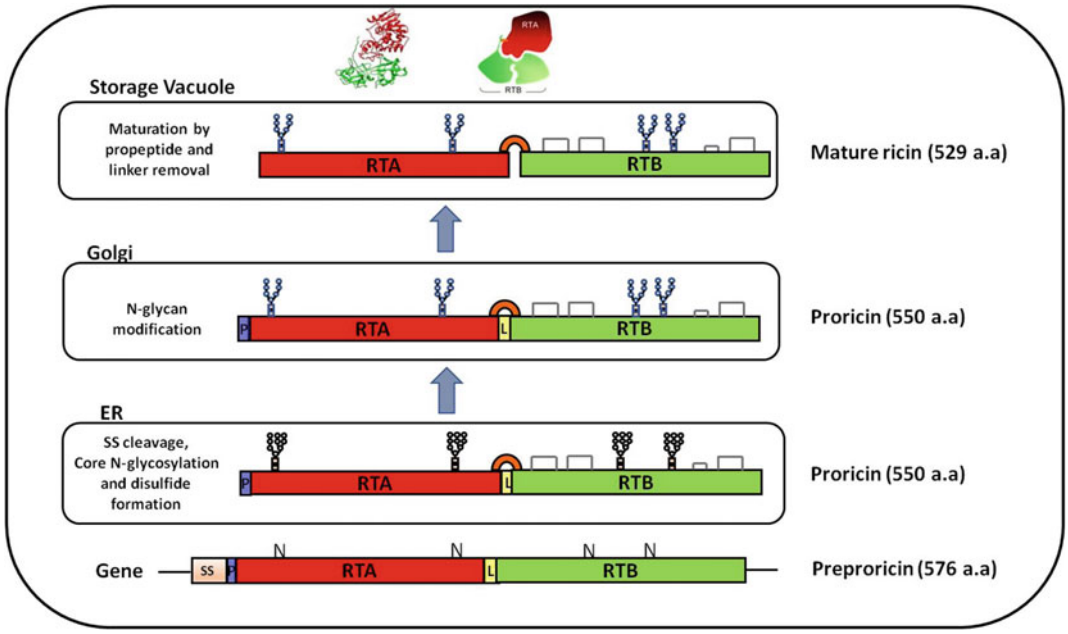

Fig. 12.1 Biogenesis of ricin and intracellular trafficking of ricin and its precursors in *Ricinus* endosperm cells. The process happens in three phases: (i) co-translational import into, and modification of ricin in the endoplasmic reticulum (ER), (ii) export into the Golgi complex and further modifications in glycosylation and (iii) export into the protein bodies, maturation and storage. Preproricin with 576 amino acid (a.a.) residues is synthesized with SS, signal sequence (26 a.a.); P, propeptide (9 a.a.); RTA, A-chain of ricin (267 a.a.); L, linker peptide (12 a.a.); RTB, B-chain of ricin (262 a.a.) and four sites of glycosylation (N). Preproricin is co-translationally translocated into Endoplasmic reticulum (ER), and the signal sequence is cleaved to give proricin (550 a.a.). In ER, along with N-glycosylation (indicated as attachments to the protein in place of N), disulphide bridges are formed; orange bracket—interchain disulphide and grey brackets—intrachain disulphides. In Golgi complex, glycan modification takes place and then ricin is transported into storage vacuoles or protein bodies, where the propeptide and linker peptides are removed coupled with conformational changes. Also shown is the ricin X-ray structure and a cartoon showing the arrangement of the chains and the position of the interchain disulphide bond. Redrawn based on Lord and Spooner (2011). Toxins (Basel), 3:787–801 an open access article under Creative Commons Attribute Licence, http://creativecommons.org/licenses/by/3.0/

by Golgi-located fucosyl transferase (Zhang and Staehelin 1992) that renders the oligosaccharide resistant to peptide: N-glycanase F (Di Cola et al. 2005).

(iii). **Export into the protein bodies, maturation and storage:** Proricin is then transported to the protein storage vacuole (PSV) through the secretory pathway, and it is processed to its mature disulphide-linked heterodimeric form by removal of both the N-terminal 9-residue propeptide and the 12-amino acid internal linker propeptide (Harley and Lord 1985). The enzyme responsible is likely to be a cysteine proteinase known as vacuolar processing enzyme (VPE) (Hara-Nishimura et al. 1993) because it cleaves at the C-terminal side of asparagine residues and both proricin propeptides terminate in asparagine. This type of proteolysis is typical in the maturation of seed storage proteins (Hara-Nishimura et al. 1991, 1995; Hiraiwa et al. 1997; Shimada et al. 2003; Wang et al. 2009), although VPE has also been identified as a caspase involved in vacuole-mediated cell death in vegetative tissues (Hatsugai et al. 2004; Hara-Nishimura et al. 2005). In the seed,

VPE is transported in a latent form to vacuoles, where it becomes activated (Hara-Nishimura et al. 1993). Despite the generation of mature ricin throughout the seed maturation process, there are no signs of endogenous ribosome damage. This strongly indicates that reduction of the interchain disulphide bond and translocation of the active RTA form the PSV to the cytosol does not occur. Thus, the generation of the active ricin holotoxin only when sheltered in vacuoles is yet another safeguard that permits castor bean seeds to store large quantities of a highly potent toxin without putting its own survival at risk.

Ricin biosynthesis in heterologous system: The biosynthetic events in castor bean endosperm have been recapitulated in tobacco protoplasts using transient expression and metabolic labelling (Frigerio et al. 1998b). These experiments provided a first indication that the linker propeptide may contain the vacuolar sorting signal. When constructs encoding preproRTA (RTA chain along with the N-terminal sequence of 1–35 aminoacids of the native preproricin protein) and preRTB (RTB chain with the linker peptide of 12 aminoacids that separate RTA and RTB in the native ricin protein) were co-expressed, RTA and RTB assembled into disulphide-linked heterodimers which were completely secreted, rather than targeted to the vacuole (Frigerio et al. 1998b). The linker peptide was sufficient to restore vacuolar sorting of the RTA–RTB heterodimers when appended to the C-terminus of RTA or to the N-terminus of RTB (Frigerio et al. 2001). The linker peptide was also sufficient to target two reporter proteins, secreted phaseolin (Frigerio et al. 1998a) and GFP/RFP, to the vacuole (Frigerio et al. 2001; Hunter et al. 2007). The 12-residue intervening linker peptide contains the amino acid motif LLIRP, which is reminiscent of the sequence-specific vacuolar sorting signal (ssVSS) NPIRL, found in proteins such as sweet potato sporamin (Matsuoka and Nakamura 1991; Koide et al. 1997) and barley aleurain (Holwerda et al. 1992). The bulky hydrophobic side chain of isoleucine has been identified as crucial within this sorting motif (Matsuoka and Nakamura 1999). Indeed, mutagenesis of Ile to Gly in the ricin linker resulted in complete secretion of proricin (Frigerio et al. 2001) or fluorescent reporter proteins (Hunter et al. 2007), indicating that the linker is a *bona fide*, ssVSS. The linker was also the first internal propeptide to be identified that carries vacuolar sorting information (Vitale and Raikhel 1999). It was subsequently shown that the position of the ricin linker within a reporter protein was important, with the linker functioning in an isoleucine-dependent manner when positioned at the N-terminus but not at the C-terminus, thus reinforcing the idea that this was a "canonical" ssVSS (Jolliffe et al. 2003). The description of the internal ssVSS of proricin was followed by the discovery that propeptide II of castor bean pro2S albumin (pro2SA) also contains a ssVSS (Brown et al. 2003). This finding also indicated that leucine can perform the same function of isoleucine within the VSS. Analysis of the *Arabidopsis* vacuolar proteome has revealed that the consensus sequence for ssVSS is indeed less strict than initially anticipated on the basis of a handful of model proteins containing NPIRL and allows for the presence of leucine as well as isoleucine (Carter et al. 2004).

In earlier years, the working model for vacuolar sorting hypothesized the existence of separate vacuoles within plant cells: a vegetative, lytic vacuole and a PSV (Paris et al. 1996). Accordingly, at least two vacuolar sorting pathways were identified (Matsuoka et al. 1995): one mediated by ssVSS and directed to the lytic vacuole, and one mediated by C-terminal, hydrophobic VSS and directed to the PSV (Matsuoka and Neuhaus 1999; Vitale and Raikhel 1999; Vitale and Hinz 2005). Therefore, the model predicted that all proteins carrying a ssVSS would be recognized by a trans-Golgi-localized vacuolar sorting receptor (first named BP-80, then renamed VSR (Paris and Neuhaus 2002)) and delivered to the lytic vacuole. As the VSR was first identified in clathrin-coated vesicle

(CCV)-enriched preparations (Kirsch et al. 1994; Ahmed et al. 1997), the hypothesis was that VSS-carrying proteins would be packaged into CCV and delivered to the lytic vacuole. Proteins directed to the PSV were hypothesized to follow a distinct route, possibly involving aggregation (Castelli and Vitale 2005), and exit from early Golgi stacks (Hillmer et al. 2001), but no interaction with VSR (Robinson et al. 2005; Vitale and Hinz 2005). Therefore, the discovery that two castor bean proteins, proricin and pro2SA, carry ssVSS and yet are targeted to the PSV challenged this model and raised the question as to whether these proteins would interact with a VSR-like receptor. Affinity chromatography experiments using immobilized ricin linker peptides identified castor bean proteins cross-reacting with *Arabidopsis* VSR antibodies in an exquisitely sequence-specific manner, thus strongly suggesting that two storage proteins are indeed VSR ligands in maturing castor bean endosperm (Jolliffe et al. 2004). Sub-cellular fractionation also revealed that proricin co-fractionated with VSR and clathrin heavy chain (Jolliffe et al. 2004). More recently, a secretory reporter (red fluorescent protein) carrying the ricin linker VSS (Hunter et al. 2007) was found to be mis-sorted to the apoplast in mutant *Arabidopsis* seeds lacking the most abundant VSR isoform, VSR1 (Craddock et al. 2008). Taken together, these observations indicate that ricin is a *bona fide* VSR substrate. The wider implication of these findings is that VSR may be central to the sorting of the majority of vacuolar proteins, including storage proteins. In addition, the fact that the existence of multiple vacuoles within plant cells seems to be the exception rather than the rule (Frigerio et al. 2008) and the recent finding that proteins carrying different types of VSS converge at the pre-vacuolar compartment (Miao et al. 2008) seems to indicate that multiple transport routes may serve the same vacuolar destination.

Though no information is available on the fate of mis-folded proricin, headway has been made with respect to the fate of RTA when expressed independently without RTB, wherein the RTA is targeted for degradation in the cell (Di cola et al. 2005). This, though less relevant to proricin (which includes both RTA and RTB), enabled the discovery of the operation of endoplasmic reticulum associated degradation (ERAD) (Sommer and Jentsch 1993; Jensen et al. 1995; Ward et al. 1995; Wiertz et al. 1996), the discussion of which is beyond the scope of this review. Also, it has been demonstrated that RTB targeted to ER in tobacco protoplast is targeted for degradation in a vacuole-independent mechanism (Chamberlain et al. 2008).

12.6 Biogenesis of RCA

The signal peptide and two propeptides in the RCA precursor are identical to the equivalent sequences in preproricin (Roberts et al. 1985), and the transport mechanism is essentially the same for both lectins (Butterworth and Lord 1983). However, the assembly of RCA into a disulphide bonded tetramer has not been studied until recently. It appears that single-chain proRCA precursors are folded and glycosylated in the ER such that they are competent for transport without assembly into precursor dimers. Consequently, they are delivered to vacuoles along a Brefeldin A (BFA)-sensitive pathway that involves trafficking through the Golgi and fucosylation of one of the RCA B-chain glycans (Lord 1985a, b). Upon arrival in the vacuole, proRCA molecules are proteolytically processed to remove the propeptides. Surprisingly, the assembly into B–A–A–B tetramers takes place within PSV rather gradually, leading eventually to a mixed population of B–A and B–A–A–B forms of RCA in the mature endosperm. This suggests that the ER may not be the only site for disulphide bond formation in plant cells (Marshall et al. 2010).

12.7 Detection of Ricin and RCA

It has been well established, through X-ray crystallography and other biophysical methods, that ricin gains cellular entry through the lectin-binding properties associated with RTB, which preferentially binds to the abundant

galactose-containing glycoproteins and glycolipids that line the cell surface. Ricin bound to the cell surface is internalized via endocytic vesicles which facilitate its retrograde transport through the Golgi and endoplasmic reticulum, after which it is extruded into the cytosol where the A-chain (**RTA**) of ricin inactivates ribosomes with high efficiency by deadenylating the A residue at 4324 position in 28SrRNA. Intracellular uptake of ricin, its retrograde transport in the mammalian cell and ribosome inactivation by RTA is depicted in Fig. 12.2.

Many different approaches towards developing simple, reliable and sensitive methods for ricin detection have been investigated (reviewed by Bozza et al. 2015). Based on the type of contaminations expected, it is critical that such methods be capable of distinguishing the active and inactive forms of ricin. There are three conditions necessary for a ricin molecule to be active and effective ribosome-inactivating protein —presence of both RTA and RTB, presence of active site in RTB to bind the galactoside residue on cell membrane and presence of the active site in RTA to bind and deadenylate the specific adenine residue in 28SrRNA. Therefore, detection methods must address these principles if distinction between active and inactive forms of ricin is to be made. When there is an intentional attack using ricin, to help in site decontamination and sample disposal, assays that can detect biologically active ricin will be needed to ascertain the level of effective decontamination of the site. Such assays would also help in developing therapeutic regimens at the site of contamination (Bozza et al. 2015). Ricin detection methods have been developed and validated in relation to the biological steps that occur during ricin intoxication; presence of ricin protein in the contaminated sample or area can be detected by immuno-based assays and mass spectrometry while the contaminating castor genomic DNA or DNA specific to ricin gene could be detected using qPCR techniques, enzymatic properties of RTA could be detected by assays that monitor the deadenylation step and concomitant release of adenine moiety as well as inhibition of protein translation, and the effect of ricin in causing cell death and apoptosis is usually detected using mouse bioassays and cell-based cytotoxicity assays which can detect ricin triggered cell death (for specific references, see the review by Bozza et al. (2015)).

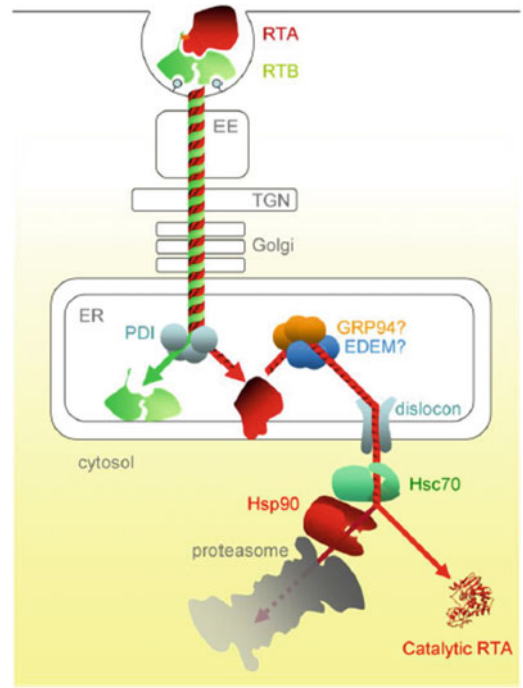

Fig. 12.2 Route ricin takes in mammalian cells and its cytotoxic effect. After entering the cell through either clathrin-dependent or clathrin-independent endocytosis, ricin traffics via vesicular carriers through the early endosomes (EE) and the trans-Golgi network (TGN) to the ER. ER processing events include separation of RTA and RTB, interaction of RTA with the ER membrane and likely interactions with luminal chaperones prior to dislocation. Post-dislocation triage by cytosolic chaperones is thought to enable a proportion of the dislocated RTA to refold. PDL—protein disulphide isomerase; EDEM—ER degradation-enhancing alpha-mannosidase I-like protein; GRP94, 94 kDa glucose-regulated protein. Reproduced from Lord and Spooner (2011). Toxins (Basel), 3:787–801 an open access article under Creative Commons Attribute Licence, http://creativecommons.org/licenses/by/3.0/

Considering the caveats in detection methods, in their review, Bozza et al. (2015) have outlined a clear perspective for the ideal detection system that would meet the requirements of monitoring and containing the harmful effects of ricin contamination. They have opined that many of the

biological assays that can detect ricin toxicity have limitations in selectivity and cannot distinguish ricin from other harmful toxins with similar mechanism of action. Therefore, it has been suggested that it is necessary to utilize an integrated approach in the development of an ideal ricin detection method with the optimal assay design that would have a rapid and efficient enrichment step, RTA activity checkpoint and a selectivity step that can distinguish ricin from other bioactive toxins. As detection of ricin, without confirming the presence of an active RTB with the lectin-binding domain does not ensure that the ricin protein is capable of penetrating a host cell membrane and exert toxicity, a viable approach for sample enrichment should be able to use sugar-conjugated materials that exploit the lectin-binding properties of the RTB. This step should be followed by detection of ricin peptide fingerprints as well as detect the in vitro released RNA substrates that have been deadenylated by active RTB through mass spectrometry methods. Additionally, it has been suggested that a robust cell-based assay, using sensitive methods such as fluorescence- or luminescence-based molecular probes that can detect ricin cytotoxicity, can be used as a confirmatory test (Bozza et al. 2015).

In order to harmonize detection capabilities in expert laboratories, an international proficiency test was organized that aimed at identifying good analytical practices (qualitative measurements) and determining a consensus concentration on a highly pure ricin reference material (quantitative measurements). Based on these studies, detection methods have been identified (Worbs et al. 2017) as adoptable in different laboratories across the world for both qualitative and quantitative measurements of ricin.

Because of the importance of ricin as a potential molecule that could be used in bio-warfare and possible environmental contamination, several new detection methods have been recently developed: rapid on-site visual detection of active ricin using a combination of highly efficient dual recognition affinity magnetic enrichment and specific gold nanoparticle probe (Sun et al. 2017), electrochemical DNA (E-DNA) biosensor-enabled detection (Daniel et al. 2017), a graphene oxide-based detection of RTB (Li et al. 2017), detection of ricin using gold nanoclusters functionalized with chicken egg white proteins as sensing probes (Selvaprakash and Chen 2017), quantitative colorimetric ricin sensing based on nanopin-cavity resonator (Fan et al. 2017), analysis of active ricin using matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (Wang et al. 2016), mass spectrometry-based rapid detection of ricin in serum using Cu-chelated magnetic beads (Zhao et al. 2016), simultaneous differentiation and quantification of ricin and RCA using an antibody-sandwich surface plasmon resonance sensor (Stern et al. 2016), etc. Recommended immunological assays as well as mass spectrometry-based strategies to identify ricin-containing samples have been enumerated (Simon et al. 2015; Kalb et al. 2015). Additionally, there are many detection methods developed based on proteomics approaches that have been summarized in a recent review (Duracova et al. 2018).

12.8 Genomics of Ricin

The ricin gene was the first plant ribosome-inactivating protein (RIP)-encoding gene to be cloned. cDNA clones encoding preproricin and preproagglutinin have revealed extensive homology. As is characteristic of all plant lectin genes, ricin gene does not contain any introns. Early Southern blot hybridization experiments (using a ricin cDNA probe) suggested that ricin and RCA belong to a multi-gene family of lectins consisting of eight members (Tregear and Roberts 1992).

However, the release of the castor bean draft genome sequence (Chan et al. 2010) revealed that there are 28 putative genes in the family, including potential pseudogenes or gene fragments, contained in a total of 17 sequence scaffolds. This distribution pattern suggested clustering of the ricin and RCA gene family members in castor genome with the largest cluster spanning 70 kb and including a group of

five family members interrupted by one non-ricin/RCA gene. Ten scaffolds contained only one member of ricin family while the remaining six scaffolds carried two to three members of ricin gene family. Earlier reports of fewer copy number of ricin family, based on the Southern blot analysis, could have missed on estimating the actual number of ricin family genes due to clustering of ricin/RCA genes at some locations in the genome through the process of tandem duplications (Chan et al. 2010).

The ricin gene uses more than one transcription start site; the most common 5′ position determined occurring 60 bp upstream from the ATG codon. There is a TATTAA sequence at −22 (with respect to the most 5′ putative transcription start point), which might function as a TATA box. A CAAGT element representing a CAAT box analogue is also present. The AGGA box is represented by ATTGA motif, but there is no evidence to support its functional role as an expression signal. RNase protection analysis was used to predict the polyadenylation site (Tregear and Roberts 1992). Two putative polyadenylation signals, with sequences AATAAA and AATAAG, have been previously identified (Lord 1985a, b) in the 3′ untranslated region of a ricin cDNA and a ricin genomic clone (Halling et a1. 1985). Investigation at the transcriptional level of the expression pattern of ricin and RCA revealed that mRNA accumulates during the post-testa stages of seed development.

Considering the complex genomics of ricin and RCA gene family, Loss-Morais et al. (2013) have reexamined the castor genome and identified 18 ricin family genes, which included nine RIP I and nine RIP II types. Further, expression pattern analyses of these identified genes using RT-qPCR at different stages of endosperm development have revealed that, in addition to the earlier identified ricin and agglutinin genes, four other RIP I and four RIP II genes are expressed during seed development in castor bean albeit with lower levels of expression (Loss-Morais et al. 2013). Thus, deeper analyses of ricin and RCA gene families are throwing newer insights into their expression pattern and this might point to their hitherto unidentified functional role in castor plant. Interestingly, in a recent review, Zhu et al. (2018) have summed up the roles ribosome-inactivating proteins play in defence responses of plants and it needs to be tested whether ricin, RCA and other RIPs in castor have definite roles in defence against pathogens and insect pests.

12.9 Approaches to Eliminate/Reduce Ricin and RCA

Considering the nutritive status of the castor de-oiled meal, especially as an animal feed, it is necessary to eliminate completely or reduce ricin and RCA content to the safe levels. The most sensible approach to reduce the toxic protein is targeting the gene coding for: (i) toxic protein, (ii) a component of the specific machinery/pathway responsible for its production and/or accumulation of the protein, (iii) a regulator of the expression of the protein either directly or indirectly, or (iv) protein that activates the inactive form of toxic protein into active toxin (not rendering the protein to become an effective toxin). Other strategies are the pharmacological or physical inactivation of the protein or stimulation of its degradation. Among these, inactivation at the gene/mRNA level is sustainable as it brings about a permanent solution. Owing to the complexity (such as multiple copies) of genomics of ricin/RCA gene family, harnessing induced mutations (using either physical agents such as X-ray and gamma ray; chemical agents such as ethyl methane sulphonate and methyl methane sulphonate; biological agents such as transposons, T-DNA) or searching for the mutated alleles from germplasm collection for ricin family genes seems to be a daunting task. Also, these processes are very slow and non-targeted. In such cases, approaches like RNAi or antisense are more effective. The most effective way to reduce toxic proteins encoded by a gene family would be to silence the genes encoding them using PTGS approaches. In gene disruption approaches, the target sequence is mutated to eliminate either gene expression or

function, whereas dominant gene silencing methods either destruct the gene transcript or prevent its synthesis. The advantages of dominant gene silencing methodologies over gene disruption approaches are twofold. First, dominant gene silencing is easier to manipulate genetically and to screen for subsequent transgenic plants. Second, dominant gene silencing can be manipulated in a spatial and temporal manner, using specific promoters.

In this section, we provide a brief summary of the attempts made so far to detoxify ricin and RCA using different methods and then outline the approaches and strategies that could be employed for bio-detoxification of ricin and RCA.

12.10 Detoxification of Ricin and RCA

12.10.1 Physical and Chemical Methods

Several attempts have been made to detoxify castor meal by destroying ricin and RCA using a number of physical and chemical methods. The physical methods included soaking in water (Anandan et al. 2005), steaming (Anandan et al. 2005), boiling (Anandan et al. 2005; Barnes et al. 2009), autoclaving (Anandan et al. 2005; Barnes et al. 2009) and heating (Anandan et al. 2005) of the oilcake. The chemical methods consisted of treatment of the cake with ammonia (Anandan et al. 2005), formaldehyde (Anandan et al. 2005), lime (Anandan et al. 2005), sodium chloride (Anandan et al. 2005), tannic acid (Anandan et al. 2005), sodium hydroxide (Anandan et al. 2005), calcium hydroxide (Barnes et al. 2009), urea (Barnes et al. 2009) and guanidine (Barnes et al. 2009). Autoclaving and lime treatment showed to be the most effective methods to destroy the toxins. However, a 60-min autoclave detoxification is very energy-demanding, making it economically unfeasible when considering large amounts of oilcake. Solvent extraction alone is less desirable as it would leave the ricin as a functional toxin within the meal unless solvents were heated to drive them from the meal to be recycled (Barnes et al. 2009). Treatment of the meal with urea and guanidine showed no substantial reduction in the presence of ricin as visualized by an antibody reaction (Barnes et al. 2009).

In another approach, it has been suggested that the ricin and RCA proteins could be inactivated by use of sal (*Shorea robusta*) seed meal, which contains very high levels of tannins. However, the treated meal shows reduced protein content and substantial levels of tannins (Gandhi et al. 1994), which non-selectively bind to proteins, destroying their enzymatic activities and reducing net protein availability of the treated castor oilmeal.

In an industrial setting involving larger quantities of castor meal, care should be taken to ensure every part of the seed or seed product was exposed to the treatment. Furthermore, because of the harsh nature of the above-described methods to destroy ricin and RCA, they can also destroy useful proteins or impair their digestibility by animals, as their digestive enzymes are also complexed by tannins. Apart from these disadvantages, all the detoxifying methods need additional knowledge and inputs from the end-user and also the extent of detoxification might vary subject to the method adopted and followed, thus always having a fair amount of risk on farmers' side to use "detoxified" oil meal as safe animal feed.

In a recent review, Akande et al. (2016) have provided a complete account of the different methods followed for detoxification of ricin and RCA in the de-oiled castor meal and its utility as animal feed.

12.10.2 Conventional Genetics

Efforts to develop lines with reduced levels of ricin and RCA in castor bean through conventional breeding strategies have not been successful. This has been ascribed to the non-availability of sufficient genetic variability (*R. communis* is a monotypic species), the existence of multiple copies of ricin and RCA genes within the genome, and the difficulty in estimating ricin concentration to select of low-ricin

plant lines. Due to these problems, some breeding trials done in the USA have not led to the identification of nil-ricin stable lines in an F6 generation (Pinkerton et al. 1999; Auld et al. 2001). However, advanced generation lines (F6) with 70–75% reduction in ricin and RCA have been developed (Auld et al. 2001). This approach is slow and costly as it requires several generations and considerable effort to develop elite lines with the desired phenotype.

12.10.3 Microbiological Detoxification

Solid-state fermentation (SSF) of castor cake has been carried out to achieve ricin detoxification and reduce allergenic potential. The fungus *Penicillium simplicissimum* was able to reduce the ricin content to non-detectable levels in addition to diminishing castor bean waste allergenic potential by approximately 16% (Godoy et al. 2009). The detoxification of oilcake by *P. simplicissimum* was evaluated during 72-h fermentation that showed a gradual reduction in ricin over time. A reduction in the amount of ricin could be observed at 24 h of fermentation and after 48 h, the ricin bands were no longer visible on SDS-PAGE, most likely because the fungus would have used ricin as nutrient source. The SSF was able to reduce, though not completely eliminate, the ricin content of the oilmeal, and also its allergenic potential, thus proving to be an effective method for residue detoxification. This is a fine example of new low-cost methodologies aimed at agro-industrial waste detoxification.

12.10.4 Bio-Detoxification Through Silencing of Ricin and RCA Gene Using Post-Transcriptional Gene Silencing (PTGS) Technologies

Gene silencing approaches have revolutionized the pursuit of suppressing the expression of target genes and have been the choice of plant biologists to eliminate the unwanted gene products in plants. If the silencing of genes is effected after the transcripts are made, such an approach is designated post-transcriptional gene silencing (PTGS). Considering the importance and suitability of this technique in eliminating ricin and RCA proteins, a brief perspective regarding PTGS concept has been provided here and the feasibility of using PTGS technology for detoxification of ricin and RCA has been provided later.

In plants, PTGS has been induced either through the use of DNA construct encoding dsRNA (double stranded RNA) or co-suppression constructs producing siRNA (small interfering RNA) or by introducing DNA constructs, which would lead to the production of antisense RNA. One of the earlier classical examples of inactivation of a gene product by PTGS technology has been the reduction of polygalacturonase enzyme to delay the onset of ripening in tomatoes where the levels of the protein were reduced up to 99% (Gray et al. 1992) through the antisense technology. Later, PTGS has been induced in plants using other strategies like co-suppression of the endogenous gene by introducing sense copies of the target gene(s). However, the relatively low frequency of PTGS achieved with antisense and co-suppression (where the endogenous gene as well as the externally introduced transgenes will be inactivated due to homology-mediated silencing of the genes leading to suppressed production of transcripts or degradation of the transcripts) requires that large populations of transgenic plants be produced in order to obtain an acceptable number of transgenic lines exhibiting sufficient degrees of target gene suppression. This can present a major limitation, particularly in species such as castor bean that have low transformation and regeneration frequency. To overcome these limitations, alternative strategies that are more efficient in achieving gene silencing have been developed in plants. Gene constructs encoding intron-spliced RNA with a hairpin structure have been shown to induce PTGS with almost 100% efficiency when directed against viruses or endogenous genes

(Smith et al. 2000). When the functional intron was replaced with any other spacer sequence or non-functional intron, then the efficiency of gene silencing was reduced significantly. In another study, Wesley et al. (2001) compared the efficiency of different types of constructs for gene silencing and have identified intron hairpin RNA strategy as the best for abolishing the gene activity. The authors have also opined that these vectors can be used for silencing whole gene families if a common conserved region among the target genes is chosen for making hpRNA construct. Recently, it has been demonstrated that if a portion of the target gene is cloned upstream of an inverted repeat of 3′ untranslated region (3′-UTR) and subsequently introduced into the plant, a high-efficient gene silencing of the target gene is achieved (Brummell et al. 2003). This type of gene silencing called silencing by heterologous 3′ untranslated regions (SHUTR) is based on the principle of transitive RNAi and has the advantage of ease and rapidity in preparation of the constructs, since a gene of interest can be inserted into a binary vector already containing the promoter and the inverted repeat of the 3′-UTR, in a single cloning step, and does not require any knowledge of the DNA sequence. The authors have demonstrated the utility of this technique for silencing many genes and transcription factors. The siRNA-mediated silencing is expected to show some off-target effects. In another milestone development, artificial microRNA approach has been adopted for targeted silencing of gene or genes of a family by taking cue from the naturally occurring microRNAs that silence genes in a very precise and regulated, both spatially and temporally, manner (Ossowski et al. 2008).

RNA silencing has the potential to simultaneously alter expression of all members of a multi-gene family in a tissue-specific manner with little collateral change in the plant. PTGS has been used in efforts to remove allergens from rice, soybean, apple, tomato and peanut. The first attempts to eliminate plant allergens were in rice (Tada et al. 1996) and soybean (Herman et al. 2003) where seed-specific promoters drove antisense or sense constructs, respectively. Although the rice transgenics and their progeny had substantially reduced levels of the targeted 14–16 kDa allergens, none of them were allergen-free probably due to insufficient sequence homology between the antisense constructs and members of the multi-gene family that encoded these allergens (Tada et al. 1996). By contrast, complete knock-down of the soybean allergen, GlymBd 30 K also known as P34, was obtained in one co-suppression line (Herman et al. 2003). Significantly, P34 elimination was maintained over three generations, and there were no observable plant morphological or reproductive differences between transgenic and wild type plants. Thus, genetic engineering can produce food crops that are hypoallergenic or are rendered harmless owing to detoxification. The advantage of this approach is that it can make foods that were once unsafe available as good sources of nutrition and calories while also paving the way for better consumer understanding and acceptance of genetically modified foods. Genetic engineering technologies have advantages over classical breeding, not only because they enlarge the range of genes and the types of mutations to be manipulated, but also due to their ability to control spatial and temporal expression patterns of the genes of interest.

A particularly interesting example of eliminating an unwanted compound through PTGS approach has been provided by Sunilkumar et al. (2006), who used tissue-specific expression of a hairpin transgene to reduce the levels of the toxic insecticidal terpenoid gossypol in cottonseed. Gossypol is toxic to humans and monogastric animals, and its accumulation in seed limits the use of cottonseed protein. However, as gossypol is a beneficial insecticidal protein elsewhere in the plant, its systemic elimination results in strongly increased susceptibility to insect attack. The enzymes δ-cadinene synthases (encoded by a gene family) are required to divert terpenoid compounds into the gossypol synthesis pathway. Sunilkumar et al. (2006) silenced these targets using a conserved hairpin expressed from the seed-specific cotton *α-globulin B* promoter. As a result, seed gossypol was reduced by up to 99%, while levels elsewhere in the plant were not

measurably changed. Other reported examples of eliminating unimportant compounds or allergens include silencing of polyphenol oxidase, to prevent enzymatic browning of potato (Wesley et al. 2001), silencing of ACC oxidase, which produces the ripening compound ethylene in tomato, resulting in altered ripening and prolonged shelf-life (Xiong et al. 2005), and silencing of the Mal d 1 family of allergenic proteins in apple (Gilissen et al. 2005).

In another important study, Allen et al. (2004) used PTGS approaches to silence the seven-member family of codeinone reductase (COR) enzymes, which catalyse the final step of morphine biosynthesis in opium poppy. The family is encoded by six highly related members, *Cor1.1–6,* and the more divergent *Cor2,* which were targeted for silencing by a chimeric hairpin sequence containing 336nt of *Cor1.1* and a 242nt sequence of *Cor2*. Specific down-regulation of *Cor* transcripts was achieved. This example clearly demonstrated that by careful selection of the homologous regions as targets while developing gene silencing vectors, a family of genes that are related by homology could be silenced effectively.

Coming to the prospects of application of PTGS approaches to eliminate ricin and RCA proteins, it involves development of transgenic castor bean plants with silencing constructs. Therefore, the basic requirements for using this approach involve availability of a good regeneration and transformation protocol and development of appropriate gene constructs to effect silencing. Castor bean is considered recalcitrant for in vitro manipulations and therefore, difficult to transform (Sujatha et al. 2009; see Chap. 14). However, there are successful examples of developing castor bean transgenics using meristem-based transformation at the Indian Council of Agricultural Research—Indian Institute of Oilseeds Research, (formerly Directorate of Oilseeds Research) Hyderabad, India (Sujatha and Sailaja 2005; Sailaja et al. 2008; Sujatha et al. 2009). Multiple genes sharing stretches of identical sequence can be effectively silenced with a single RNAi construct. This feature is especially important in species like castor bean, which has multiple copies of ricin and RCA genes. These genes are also very tightly regulated at the transcriptional level with their expression clearly restricted to the developing endosperm (Tregear and Roberts 1992; Lord et al. 1994; Frigerio and Roberts 1998; Pinkerton et al. 1999). Thus, in order to post-transcriptionally silence them, the siRNA molecules must be expressed in the same cells at the same time. Unfortunately, the commonly used constitutive CaMV35S (cauliflower mosaic virus 35S) promoter does not work well in the endosperm (Benfey and Chua 1990), and an endosperm-specific promoter should be used.

In collaboration with ARS-USDA, Texas Tech University scientists have developed a transformation protocol in castor bean and they are trying to develop transgenic castor bean with low or no ricin and RCA content using antisense technology (Auld et al. 2001). The major objective of castor bean genetic transformation in the USA is to develop ricin-free castor bean varieties with the potential to create new economic opportunities for farmers and processors. The problem is tackled through a PTGS approach (Auld et al. 2001). At Arcadia Biosciences Inc, Canada efforts are on to develop ricin-free castor bean by adopting targeting induced local lesions in genomes (TILLING) technology. In this approach, single-nucleotide polymorphisms in the ricin gene(s) are identified to be later brought together into one genetic background (www.arcadiabio.com). The most obvious advantage of TILLING for silencing toxic substances is that it is a knockout approach, while RNAi is a knock-down approach. Also, TILLING generates "non-GMO" (non-genetically modified organism) or non-transgenic plants. However, the high copy number of ricin genes in the castor bean genome may prevent the use of TILLING to eliminate all copies of the toxin genes.

In a recent development, a Brazilian research group based in Embrapa resorted to RNAi (intron hairpin)-mediated silencing of ricin and RCA genes in castor bean (Sousa et al. 2017). Non-detection of ricin protein in transgenic castor bean lines, lack of hemagglutination activity and non-toxicity of the de-oiled meal from

transgenic lines further established the effective silencing of ricin and RCA mediated by the intron hairpin RNAi strategy. This has ushered in a new era of utilizing the detoxified, protein-rich, de-oiled meal as a good animal feed.

12.11 Conclusion

The castor oilcake is a promising candidate for animal feed. The toxic endosperm proteins, ricin and RCA, severely limit the utility of castor cake after oil extraction. The total elimination or inactivation of these toxic proteins from castor cake is extremely important before it can be used as an animal feed. The global shortage of protein sources for animal feed has prompted a search for alternative sources. Attempts have been made to detoxify castor cake using physical and chemical approaches, conventional genetics, microbiological detoxification and biotechnological interventions like PTGS and TILLING. Future attempts could exploit the cell biology of ricin and RCA.

It is interesting to note that the cultivation of castor bean was abandoned in the Texas region of the USA to avoid health hazards associated with the accidental ingestion of castor bean seeds by humans or animals. However, McKeon and Chen (2001), with the ultimate goal of realizing ricin-free transgenic castor bean plants, have reported the successful development of castor bean transgenic lines using agro-infection method, in which the flower buds were wounded and infiltrated with *Agrobacterium* culture, seeds were allowed to develop and the resultant plants from such seeds were tested for transgenic nature. Interestingly, the title of the paper "High-tech castor bean plants may open door to domestic production" is an indication of the health hazard caused by ricin. Among the several options available to down-regulate multiple copies of highly homologous toxin genes, PTGS technology-based strategies are very promising, as has been already proven in different examples. Ricin and RCA proteins being encoded by a family of highly homologous genes, this should be a clear case for bio-detoxification approach through development of appropriate transgenic lines by employing efficient PTGS strategies. The basis requirement for using transgenic approach is the availability of a repeatable transformation protocol, and therefore, concerted efforts are needed in this direction in castor bean. Recent report (Sousa et al. 2017) demonstrates the feasibility of developing ricin- and RCA-free castor bean plants.

Complete exploitation of the potential of PTGS in crop improvement will be possible with the development of more specific temporal/spatially regulated silencing vectors. In author's laboratory, a set of gene silencing constructs that use ihpRNAi, transitive RNAi and artificial microRNA principles have been developed by targeting the sequences that are common between ricin and RCA genes and by using the promoter of native ricin gene. These constructs will be used to transform castor bean and assess the efficiency of each of the approaches. Multipronged approaches combining PTGS with conventional genetics or other methods may, for instance, further reduce or completely eliminate ricin and RCA in the USDA Texas castor bean lines with reduced ricin and RCA. It is also essential to understand the interaction and the network of various macromolecular complexes and their modulators involved in the phenomenon of PTGS to provide a better platform to harness the potential of this natural cellular process for targeted gene silencing. Additionally, with the current swift advancement of plant genomics, proteomics and metabolomics, there will be both a store of novel genes to evaluate and improve genetic engineering strategies for continued food crop improvement and therefore, the possibility of ricin- and RCA-free castor bean could be a reality.

Acknowledgements The financial assistance provided by the Department of Science and Technology, Government of India, in the form of research grant (SR/SO/BB-58/2005/DST) for the work which was carried out at the Indian Institute of Oilseeds Research (Formerly Directorate of Oilseeds Research), Hyderabad, India, is duly acknowledged. Authors also acknowledge the support extended by the Director, ICAR-IIOR, during the course of investigation and preparation of the manuscript.

References

Ahmed SU, Bar-Peled M, Raikhel NV (1997) Cloning and subcellular location of an *Arabidopsis* receptor-like protein that shares common features with protein-sorting receptors of eukaryotic cells. Plant Physiol 114:325–336

Allen RS, Millgate AG, Chitty A, Thisleton Miller AC, Fist A, Gerlach WL, Larkin P (2004) RNAi-mediated replacement of morphine with the non narcotic alkaloid reticuline in opium poppy. Nat Biotechnol 22:1559–1566

Akande TO, Odunsi AA, Akinfala EO (2016) A review of nutritional and toxicological implications of castor bean (*Ricinus communis* L.) meal in animal feeding systems. J Anim Physiol Anim Nutr 100:201–210

Anandan S, Anil Kumar GK, Ghosh J, Ramachandra KS (2005) Effect of different physical and chemical treatments on detoxification of ricin in castor cake. Anim Feed Sci Technol 120:159–168

Argent RH, Roberts LM, Wales R, Robertus JD, Lord JM (1994) Introduction of a disulfide bond into ricin A chain decreases the cytotoxicity of the ricin holotoxin. J Biol Chem 269:26705–26710

Auld DL, Rolfe RD, McKeon TA (2001) Development of castor with reduced toxicity. J New Seeds 3:61–69

Barnes DJ, Baldwin BS, Braasch DA (2009) Ricin accumulation and degradation during castor seed development and late germination. Ind Crops Prod https://doi.org/10.1016/j.indcrop.2009.04.003

Benfey PN, Chua NH (1990) The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants. Science 250:959–966

Bozza WP, Tolleson WH, Rosado LA, Zhang B (2015) Ricin detection: tracking active toxin. Biotechnol Adv 33(1):117–123. https://doi.org/10.1016/j.biotechadv.2014.11.012

Brown JC, Jolliffe NA, Frigerio L, Roberts LM (2003) Sequence-specific, golgi-dependent targeting of the castor bean 2S albumin to the vacuole in tobacco protoplasts. Plant J 36:711–719

Brummell DA, Balint-Kurti PJ, Harpster MH, Palys JM, Oeller PW, Gutterson N (2003) Inverted repeat of a heterologous 3′-untranslated region for high-efficiency, high throughput gene silencing. Plant J 33:793–800

Butterworth AG, Lord JM (1983) Ricin and *Ricinus communis* agglutinin subunits are all derived from a single-size polypeptide precursor. Eur J Biochem 137:57–65

Carter C, Pan S, Zouhar J, Avila EL, Girke T, Raikhel NV (2004) The vegetative vacuole proteome of *Arabidopsis thaliana* reveals predicted and unexpected proteins. Plant Cell 16:3285–3303

Castelli S, Vitale A (2005) The phaseolin vacuolar sorting signal promotes transient, strong membrane association and aggregation of the bean storage protein in transgenic tobacco. J Exp Bot 56:1379–1387

Cawley DB, Houston LL (1979) Effect of sulfhydryl reagents and protease inhibitors on sodium dodecyl sulfate-heat induced dissociation of *Ricinus communis* agglutinin. Biochim Biophys Acta 581:51–62

Cawley DB, Hedblom ML, Hoffman EJ, Houston LL (1977) Differential ricin sensitivity of rat liver and wheat germ ribosomes in polyuridylic acid translation. Arch Biochem Biophys 182:690–695

Chamberlain KL, Marshall RS, Jolliffe NA, Frigerio L, Ceriotti A, Lord JM, Roberts LM (2008) Ricin B chain targeted to the endoplasmic reticulum of tobacco protoplasts is degraded by a CDC48- and vacuole-independent mechanism. J Biol Chem 283:33276–33286

Chan AP, Crabtree T, Zhao Q, Lorenzi H, Orvis J, Puiu D, Melake-Berhan A, Jones KM, Redman J, Chen G, Cahoon EG, Gedil M, Stanke M, Haas J, Wortman JR, Fraser-Liggett CM, Ravel J, Rabiowicz PD (2010) Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotechnol. Online 22 Aug 2010. https://doi.org/10.1038/nbt.1674

Chen GQ, He X, McKeon TA (2005) A simple and sensitive assay for distinguishing the expression of ricin and *Ricinus communis* agglutinin genes in developing castor seed (*Ricinus communis* L.). J Agri Food Chem 53:2358–2361

Craddock C, Hunter P, Szakacs E, Hinz G, Robinson D, Frigerio L (2008) Lack of a vacuolar sorting receptor leads to non-specific missorting of soluble vacuolar proteins in *Arabidopsis* seeds. Traffic 9:408–416

Daniel J, Fetter L, Jett S, Rowland TJ, Bonham AJ (2017) Electrochemical aptamer scaffold biosensors for detection of botulism and ricin proteins. Meth Mol Biol 1600:9–23. https://doi.org/10.1007/978-1-4939-6958-6_2

Di Cola A, Frigerio L, Lord JM, Roberts LM, Ceriotti A (2005) Endoplasmic reticulum associated degradation of ricin A chain has unique and plant-specific features. Plant Physiol 137:287–296

Duracova M, Klimentova J, Fucikova A, Dresler J (2018) Proteomic methods of detection and quantification of protein toxins. Toxins (Basel) 10(3):99. https://doi.org/10.3390/toxins10030099

Endo Y, Tsurugi K (1987) RNA N-glycosidase activity of ricin A-chain: mechanism of action of the toxin lectin ricin in eukaryotic ribosomes. J Biol Chem 262:8128–8130

Fan JR, Zhu J, Wu WG, Huang Y (2017, January) Plasmonic metasurfaces based on nanopin-cavity resonator for quantitative colorimetric ricin sensing. Small 13(1). https://doi.org/10.1002/smll.201601710

Ferrini JB, Martin M, Taupiac MP, Beaumelle B (1995) Expression of functional ricin B chain using the baculovirus system. Plant Physiol 233:772–777

Frigerio L, Roberts LM (1998) The enemy within: ricin and plant cells. J Exp Bot 49:1473–1480

Frigerio L, de Virgilio M, Prada A, Faoro F, Vitale A (1998a) Sorting of phaseolin to the vacuole is

saturable and requires a short C-terminal peptide. Plant Cell 10:1031–1042

Frigerio L, Hin G, Robinson DG (2008) Multiple vacuoles in plant cells: rule or exception? Traffic 9:1564–1570

Frigerio L, Jolliffe NA, Di Cola A, Felipe HD, Paris N, Neuhaus JM, Lord JM, Ceriotti A, Roberts LM (2001) The internal propeptide of the ricin precursor carries a sequence specific determinant for vacuolar sorting. Plant Physiol 126:167–175

Frigerio L, Vitale A, Lord JM, Ceriottim A, Roberts LM (1998b) Free ricin A chain, proricin and native toxin have different cellular fates when expressed in tobacco protoplasts. J Biol Chem 273:14194–14199

Fu D, Uauy C, Blechl A, Dubcovsky J (2007) RNA interference for wheat functional gene analysis. Transgen Res 16:689–701

Gandhi VM, Cherian KM, Mulky MJ (1994) Detoxification of castor seed meal by interaction with sal seed meal. J Amer Oil Chem Soc 71:827–831

Gilissen L, Bolhaar STH, Matos CR, Rouwendal GA, Boone M, Krens EA, Zuidmeer L, van Leeuwen A, Akkerdaas Hoffmann-Sommergruber K, Knulst AC, Bosch D, van de Weg WE, van Ree R (2005) Silencing the major apple allergen Mal d 1 by using the RNA interference approach. J Allergy Clin Immunol 115:364–369

Godoy MG, Gutarra MLE, Maciel FM, Felix SP, Bevilaqua JV, Machado OLT, Freire DMG (2009) Use of a low-cost methodology for biodetoxification of castor bean waste and lipase production. Enzyme Microb Technol 44:317–322

Gray J, Picton S, Shabeer J, Schuch W, Grierson D (1992) Molecular biology of fruitripening and its manipulation with antisense genes. Plant Mol Biol 9:69–87

Halling KC, Halling AC, Murray EE, Ladin BF, Houston LL, Weaver RF (1985) Genomic cloning and characterization of a ricin gene from *Ricinus communis*. Nucl Acid Res 13:8019–8033

Hara-Nishimura I, Hatsugai N, Nakaune S, Kuroyanagi M, Nishimura M (2005) Vacuolar processing enzyme: an executor of plant cell death. Curr Opin Plant Biol 8:404–408

Hara-Nishimura I, Inoue K, Nishimura M (1991) A unique vacuolar processing enzyme responsible for conversion of several proprotein precursors into the mature forms. FEBS Lett 294:89–93

Hara-Nishimura I, Shimada T, Hiraiwa N, Nishimura M (1995) Vacuolar processing enzyme responsible for the maturation of seed proteins. J Plant Physiol 145:632–640

Hara-Nishimura I, Takeuchi Y, Nishimura M (1993) Molecular characterization of a vacuolar processing enzyme related to a putative cysteine proteinase of *Schistosoma mansoni*. Plant Cell 5:1651–1659

Harley SM, Beevers H (1982) Ricin inhibition of in vitro protein synthesis by plant ribosomes. Proc Natl Acad Sci USA 79:5935–5938

Harley SM, Lord JM (1985) In vitro endoproteolytic cleavage of castor bean lectin precursors. Plant Sc 41:111–116

Hatsugai N, Kuroyanagi M, Yamada K, Meshi T, Tsuda S, Kondo M, Nishimura M, Hara- Nishimura I (2004) A plant vacuolar protease, VPE, mediates virus-induced hypersensitive cell death. Science 305:855–858

Herman EM, Helm RM, Jung R, Kinney AJ (2003) Genetic modification removes an immunodominant allergen from soybean. Plant Physiol 132:36–43

Hillmer S, Movafeghi A, Robinson DG, Hinz G (2001) Vacuolar storage proteins are sorted in the cis-cisternae of the pea cotyledon golgi apparatus. J Cell Biol 152:41–50

Hiraiwa N, Kondo M, Nishimura M, Hara-Nishimura I (1997) An aspartic endopeptidase is involved in the breakdown of storage proteins in protein-storage vacuoles of plants. Eur J Biochem 246:133–141

Holwerda BC, Padgett HS, Rogers JC (1992) Proaleurain vacuolar targeting is mediated by short contiguous peptide interactions. Plant Cell 4:307–318

Hunter PR, Craddock CP, Di Benedetto S, Roberts LM, Frigerio L (2007) Fluorescent reporter proteins for the tonoplast and the vacuolar lumen identify a single vacuolar compartment in *Arabidopsis* cells. Plant Physiol 145:1371–1382

Jensen TJ, Loo MA, Pind S, Williams DB, Goldberg AL, Riordan JR (1995) Multiple proteolytic systems, including the proteasome, contribute to CFTR processing. Cell 83:129–135

Jolliffe NA, Brown JC, Neumann U, Vicre M, Bachi A, Hawes C, Ceriotti A, Roberts LM, Frigerio L (2004) Transport of ricin and 2S albumin precursors to the storage vacuoles of *Ricinus communis* endosperm involves the golgi and VSR-like receptors. Plant J 39:821–833

Jolliffe NA, Ceriotti A, Frigerio L, Roberts LM (2003) The position of the proricin vacuolar targeting signal is functionally important. Plant Mol Biol 51:631–641

Jolliffe NA, Di Cola A, Marsden CJ, Lord JM, Ceriotti A, Frigerio L, Roberts LM (2006) The N-terminal ricin propeptide influences the fate of ricin A-chain in tobacco protoplasts. J Biol Chem 281:23377–23385

Kalb SR, Schieltz DM, Becher F, Astot C, Fredriksson SA, Barr JR (2015) Recommended mass spectrometry-based strategies to identify ricin-containing samples. Toxins (Basel) 7:4881–4894. https://doi.org/10.3390/toxins7124854

Katzin BJ, Collins EJ, Robertus JD (1991) The structure of ricin A chain at 2.5 Å. Proteins Struct Funct Genet 10:251–259

Kirsch T, Paris N, Butler JM, Beevers L, Rogers JC (1994) Purification and initial characterization of a potential plant vacuolar targeting receptor. Proc Natl Acad Sci USA 91:3403–3407

Koide Y, Hirano H, Matsuoka K, Nakamura K (1997) The N-terminal propeptide of the precursor to

sporamin acts as a vacuole-targeting signal even at the C terminus of the mature part in tobacco cells. Plant Physiol 114:863–870

Lamb FI, Roberts LM, Lord JM (1985) Nucleotide sequence of cloned cDNA coding for preproricin. Eur J Biochem 148:265–270

Li CH, Xiao X, Tao J, Wang DM, Huang CZ, Zhen SJ (2017) A graphene oxide-based strand displacement amplification platform for ricin detection using aptamer as recognition element. Biosens Bioelec 15(91):149–154. https://doi.org/10.1016/j.bios.2016.12.010

Lord JM (1985a) Precursors of ricin and *Ricinus communis* agglutinin glycosylation and processing during synthesis and intracellular transport. Eur J Biochem 146:411–416

Lord JM (1985b) Synthesis and intracellular transport of lectin and storage protein precursors in endosperm from castor bean. Eur J Biochem 146:403–409

Lord JM, Spooner RA (2011) Ricin trafficking in plant and mammalian cells. Toxin 3:787–801

Lord JM, Roberts LM, Robertus JD (1994) Ricin: structure, mode of action and some current applications. FASEB J 8:201–208

Loss-Morais G, Turchetto-Zolet AC, Etges M, Cagliari A, Körbes AP, Maraschin MS, Margis-Pinheiro M, Margis R (2013) Analysis of castor bean ribosome-inactivating proteins and their gene expression during seed development. Genet Mol Biol 36:74–86

Marshall RS, Frigerio L, Roberts LM (2010) Disulfide formation in plant storage vacuoles permits assembly of a multimeric lectin. Biochem J 427:513–521

Matsuoka K, Nakamura K (1991) Propeptide of a precursor to a plant vacuolar protein required for vacuolar targeting. Proc Natl Acad Sci USA 88:834–838

Matsuoka K, Nakamura K (1999) Large alkyl side-chains of isoleucine and leucine in the NPIRL region constitute the core of the vacuolar sorting determinant of sporamin precursor. Plant Mol Biol 41:825–835

Matsuoka K, Neuhaus JM (1999) Cis-elements of protein transport to the plant vacuoles. J Exp Bot 50:165–174

Matsuoka K, Bassham DC, Raikhel NV, Nakamura K (1995) Different sensitivity to wortmannin of two vacuolar sorting signals indicates the presence of distinct sorting machineries in tobacco cells. J Cell Biol 130:1307–1318

McKeon TA, Chen GQ (2001) High-tech castor plants may open door to domestic production. ARS Magaz 49:12–13

Miao Y, Li KY, Li HY, Yao X, Jiang L (2008) The vacuolar transport of aleurain-GFP and 2S albumin-GFP fusions is mediated by the same pre-vacuolar compartments in tobacco BY-2 and *Arabidopsis* suspension cultured cells. Plant J 56:824–839

Montfort W, Villafranca JE, Monzingo AF, Ernst S, Katzin B, Rutenber E, Xuong NH, Hamlin R, Robertus JD (1987) The three-dimensional structure of ricin at 2.8 A. J Biol Chem 262:5398–5403

Nicolson GL, Blaustein J, Etzler ME (1974) Characterization of two plant lectins from *Ricinus communis* and their quantitative interaction with a murine lymphoma. Biochemistry 13:196–204

Ogunniyi DS (2006) Castor oil: a vital industrial raw material. Bioresour Technol 97:1086–1091

Ossowski S, Schwab R, Weigel D (2008) Gene silencing in plants using artificial microRNAs and other small RNAs. Plant J 53:674–690

Paris N, Neuhausm JM (2002) BP-80 as a vacuolar sorting receptor. Plant Mol Biol 50:903–914

Paris N, Stanley CM, Jones RL, Rogers JC (1996) Plant cells contain two functionally distinct vacuolar compartments. Cell 85:563–572

Pinkerton SD, Rolfe RD, Auld DL, Ghetie V, Lauterbach BF (1999) Selection of castor with divergent concentrations of ricin and *Ricinus communis* agglutinin. Crop Sci 39:353–357

Richardson PT, Westley M, Roberts LM, Gould JH, Colman A, Lord JM (1989) Recombinant proricin binds galactose but does not depurinate 28S ribosomal RNA. FEBS Lett 255:15–20

Roberts LM, Lord JM (1981) Protein biosynthetic capacity in the endosperm tissue of ripening castor bean seeds. Planta 152:420–427

Roberts LM, Lamb FI, Pappin DJC, Lord JM (1985) The primary sequence of *Ricinus communis* agglutinin. Comparison with ricin. J Biol Chem 260:15682–15688

Robinson DG, Oliviusson P, Hinz G (2005) Protein sorting to the storage vacuoles of plants: a critical appraisal. Traffic 6:615–625

Rutenber E, Robertus JD (1991) The structure of ricin B-Chain at 2.5 Å resolution. Proteins 10:260–269

Rutenber E, Katzin BJ, Collins EJ, Mlsna D, Ernst SE, Ready MP, Robertus JD (1991) The crystallographic refinement of ricin to 2.5 Å. Proteins Struct Funct Genet 10:240–250

Sailaja M, Tarakeswari M, Sujatha M (2008) Stable genetic transformation of castor (*Ricinus communis* L.) via particle gun-mediated gene transfer using embryo axes from mature seeds. Plant Cell Rep 27:1509–1519

Scarpa A, Guerci A (1982) Various uses of the castor oil plant (*Ricinus communis* L.) a review. J Ethnopharmacol 5:117–137

Selvaprakash K, Chen YC (2017) Detection of ricin by using gold nanoclusters functionalized with chicken egg white proteins as sensing probes. Biosens Bioelec 15(92):410–416. https://doi.org/10.1016/j.bios.2016.10.086

Shimada T, Yamada K, Kataoka M, Nakaune S, Koumoto Y, Kuroyanagi M, Tabata S, Kato T, Shinozaki K, Seki M, Kobayashi M, Kondo M, Nishimura M, Hara-Nishimura I (2003) Vacuolar processing enzymes are essential for proper processing of seed storage proteins in *Arabidopsis thaliana*. J Biol Chem 278:32292–32299

Simon S, Worbs S, Avondet MA, Tracz DM, Dano J, Schmidt L, Volland H, Dorner BG, Corbett CR (2015) Recommended immunological assays to screen for ricin-containing samples. Toxins (Basel) 7(12):4967–4986. https://doi.org/10.3390/toxins7124858

Smith NA, Singh SP, Wang M, Stoutjesdijk PA, Green AG, Waterhouse PM (2000) Total silencing by intron-spliced hairpin RNAs. Nature 407:319–320

Sommer T, Jentsch S (1993) A protein translocation defect linked to ubiquitin conjugation at the endoplasmic reticulum. Nature 365:176–179

Sousa NL, Cabral GB, Vieira PM, Baldoni AB, Aragao FJL (2017) Bio-detoxification of ricin in castor bean (*Ricinus communis* L.) seeds. Sci Rep 7:15385

Spooner RA, Watson PD, Marsden CJ, Smith DC, Moore KA, Cook JP, Lord JM, Roberts LM (2004) Protein disulphide-isomerase reduces ricin to its A and B chains in the endoplasmic reticulum. Biochem J 383:285–293

Stern D, Pauly D, Zydek M, Müller C, Avondet MA, Worbs S, Lisdat F, Dorner MB, Dorner BG (2016) Simultaneous differentiation and quantification of ricin and agglutinin by an antibody-sandwich surface plasmon resonance sensor. Biosens Bioelec 15(78):111–117. https://doi.org/10.1016/j.bios.2015.11.020

Sujatha M, Sailaja M (2005) Stable genetic transformation of castor (*Ricinus communis* L.) via *Agrobacterium tumefaciens*-mediated gene transfer using embryo axes from mature seeds. Plant Cell Rep 23:803–810

Sujatha M, Lakshminarayana M, Tarakeswari M, Singh PK, Tuli R (2009) Expression of the cry1EC gene in castor (*Ricinus communis* L.) confers field resistance to tobacco caterpillar (*Spodoptera litura Fabr*) and castor semilooper (*Achoea janata* L.). Plant Cell Rep 28:935–946

Sun J, Wang C, Shao B, Wang Z, Xue D, Liu Y, Qi K, Yang Y, Niu Y (2017, November) Fast on-site visual detection of active ricin using a combination of highly efficient dual-recognition affinity magnetic enrichment and a specific gold nanoparticle probe. Anal Chem 89 (22):12209–12216. https://doi.org/10.1021/acs.analchem.7b02944

Sunilkumar G, Campbell LM, Puckhaber L, Stipanovic RD, Rathore KS (2006) Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. Proc Natl Acad Sci USA 103:18054–18059

Sweeney EC, Tonevitsky AG, Temiakov DE, Agapov II, Saward S, Palmer RA (1997) Preliminary crystallographic characterization of ricin agglutinin. Proteins 28:586–589

Tada Y, Nakase M, Adachi T, Nakamura R, Shimada H, Takahashi M, Fujimura T, Matsuda T (1996) Reduction of 14–16 kDa allergenic proteins in transgenic rice plants by antisense gene. FEBS Lett 391:341–345

Taylor S, Massiah A, Lomonossoff G, Roberts LM, Lord JM, Hartley M (1994) Correlation between the activities of five ribosome-inactivating proteins in depurination of tobacco ribosomes and inhibition of tobacco mosaic virus infection. Plant J 5:827–835

Tregear JW, Roberts LM (1992) The lectin gene family of *Ricinus communis*: cloning of a functional ricin gene and three lectin pseudogenes. Plant Mol Biol 18:515–525

Tully RE, Beevers H (1976) Protein bodies of castor bean endosperms: isolation, fractionation and characterisation of protein components. Plant Physiol 58:710–716

Vitale A, Hinz G (2005) Sorting of proteins to storage vacuoles: how many mechanisms? Trends Plant Sci 10:316–323

Vitale A, Raikhel NV (1999) What do proteins need to reach different vacuoles? Trends Plant Sci 4:149–155

Wang D, Baudys J, Barr JR, Kalb SR (2016, July) Improved sensitivity for the qualitative and quantitative analysis of active ricin by MALDI-TOF mass spectrometry. Anal Chem 88(13):6867–6872. https://doi.org/10.1021/acs.analchem.6b01486

Wang Y, Zhu S, Liu S, Jiang L, Chen L, Ren Y, Han X, Liu F, Ji S, Liu X, Wan J (2009) The vacuolar processing enzyme OsVPE1 is required for efficient glutelin processing in rice. Plant J 58:606–617

Ward CL, Omura S, Kopito RR (1995) Degradation of CFTR by the ubiquitin-proteasome pathway. Cell 83:121–127

Wesley SV, Helliwell CA, Smith NA, Wang M, Rouse DT, Liu Q, Gooding PS, Singh SP, Abbot D, Stoutjesdijk PA, Robinson SP, Gleave AP, Green AG, Waterhouse PM (2001) Construct design for effective and high throughput gene silencing in plants. Plant J 27(6):581–590

Wiertz EJH, Jones TR, Sun L, Bogyo M, Geuze HJ, Ploegh HL (1996) The human cytomegalovirus US11 gene product dislocates MHC class I heavy chains from the endoplasmic reticulum to the cytosol. Cell 84:769–779

Wiley RG, Oeltman TN (1991) Ricin and related plant toxins: mechanisms of action and neurobiological applications. In: Keeler RF, Tu AT (eds) Handbook of natural toxins: toxicology of plant and fungal compounds, vol 6. Marcel Dekker, New York, p 665

Worbs S, Skiba M, Bender J, Zeleny R, Schimmel H, Luginbühl W, Dorner BG (2017) An international proficiency test to detect, identify and quantify ricin in complex matrices (26 Nov 2015). Toxins (Basel) 7(12): 4987–5010. https://doi.org/10.3390/toxins7124859

Xiong AS, Yao QH, Peng RH, Li X, Han PL, Fan HQ (2005) Different effects on ACC oxidase gene silencing triggered by RNA interference in transgenic tomato. Plant Cell Rep 23:639–646

Youle RJ, Huang AHC (1976) Protein bodies from the endosperm of castor bean subfractionation, protein

components, lectins and changes during germination. Plant Physiol 58:703–709

Zhang GF, Steahelin AL (1992) Functional compartmentation of the golgi apparatus of plant cells. Plant Physiol 99:1070–1083

Zhao YQ, Song J, Wang HL, Xu B, Liu F, He K, Wang N (2016) Rapid detection of ricin in serum based on Cu-chelated magnetic beads using mass spectrometry. J Amer Soc Mass Spectrom 27(4):748–751. https://doi.org/10.1007/s13361-016-1340-1

Zhu F, Zhou YK, Ji ZL, Chen XR (2018) The Plant ribosome-inactivating proteins play important roles in defense against pathogens and insect pest attacks. Front Plant Sci 9:146. https://doi.org/10.3389/fpls.2018.00146

13 Castor Bean Metabolomics: Current Knowledge and Perspectives Toward Understanding of Plant Plasticity Under Stress Condition

Paulo R. Ribeiro, Gisele A. B. Canuto, Valdinei C. Brito, Danilo L. J. Batista, Cristiane D. de Brito, Marta B. Loureiro, Daniele Takahashi, Renato Delmondez de Castro, Luzimar Gonzaga Fernandez, Henk W. M. Hilhorst and Wilco Ligterink

Abstract

Metabolomics provides vital information for the understanding of biological processes and has been vastly applied in plant studies. Several metabolite-profiling studies have correlated physiological events, such as germination or seedling establishment, with metabolic and molecular changes under different environmental conditions. Castor bean displays high plasticity during initial vegetative growth, which is reflected in the metabolome of the seeds and seedlings. In general, several metabolite-profiling techniques are required to obtain a complete response in terms of metabolism plasticity of the studied biological system. Carbohydrates, amino acids, and organic acids have been measured in castor bean seeds and seedlings by nuclear magnetic resonance, gas chromatography coupled to a quadrupole time of flight mass spectrometry (GC-TOF-MS),

P. R. Ribeiro (✉) · G. A. B. Canuto · V. C. Brito
D. L. J. Batista
Metabolomics Research Group, Instituto de Química, Universidade Federal da Bahia, Rua Barão de Jeremoabo s/n, Salvador 40170-115, Brazil
e-mail: paulodc3@gmail.com; pauloribeiro@ufba.br

G. A. B. Canuto
e-mail: gisele.canuto@ufba.br

V. C. Brito
e-mail: valbrito_10@hotmail.com

D. L. J. Batista
e-mail: dilo-luis@hotmail.com

P. R. Ribeiro · C. D. de Brito · M. B. Loureiro
D. Takahashi · R. D. de Castro · L. G. Fernandez
Laboratório de Bioquímica, Biotecnologia E Bioprodutos, Departamento de Bioquímica E Biofísica, Universidade Federal da Bahia, Reitor Miguel Calmon s/n, Salvador 40160-100, Brazil
e-mail: crisbrichta@gmail.com

M. B. Loureiro
e-mail: brunoloureiro70@gmail.com

D. Takahashi
e-mail: danitaka@hotmail.com

R. D. de Castro
e-mail: renatodel@gmail.com

L. G. Fernandez
e-mail: luzimargonzaga@gmail.com

H. W. M. Hilhorst · W. Ligterink
Wageningen Seed Lab, Laboratory of Plant Physiology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen The Netherlands
e-mail: Henk.Hilhorst@wur.nl

W. Ligterink
e-mail: wilco.ligterink@wur.nl

C. Kole and P. Rabinowicz (eds.), *The Castor Bean Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97280-0_13

as well as by high-performance liquid chromatography (HPLC). Fatty acids and some secondary metabolites have been quantified in castor bean seeds and seedlings by gas chromatography coupled to a triple-axis detector (GC-MS). In this chapter, we initially discuss how metabolomics studies suggested a possible role of gamma-aminobutyric acid (GABA) accumulation during early imbibitions and seedling establishment. Later, we consider a specific metabolic signature of castor bean: a shift in carbon–nitrogen metabolism as its main biochemical response to high temperatures. This metabolic shift is usually associated with adjusted growth, and it is likely involved in maintaining cellular homeostasis under heat stress. The castor bean metabolome has been vastly investigated, especially with regard to its ability to respond to external stimuli. These results might help us understand the molecular requirements for vigorous castor bean seed germination and seedling growth under different environmental conditions.

13.1 Introduction

13.1.1 General Aspects of Metabolomics

The concept that a complex biological system is greater than the sum of its parts is the basis of the multidisciplinary approach known as systems biology. This approach studies complex organisms and their physiology from a holistic point of view (Klassen et al. 2017). New "omics" technologies such as genomics, transcriptomics, proteomics, and metabolomics have allowed detailed and integrated analyses to understand the phenotype of an organism. Among these methodologies, metabolomics is a crucial tool to help understand biological systems dynamics, because it represents the endpoint of gene expression and enzymatic activity (Kuehnbaum and Britz-McKibbin 2013; Villas-Boas et al. 2005).

The term metabolomics was first defined as the comprehensive and quantitative analysis of all metabolites present in a given organism (Fiehn 2001). Previously, "metabonomics" was commonly used to represent "the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification" (Nicholson et al. 1999). Metabolomics now includes both concepts, and it can be defined as the qualitative and quantitative analysis of the entire set of metabolites (metabolome) expressed or modified by an organism (Canuto et al. 2015).

Metabolomics studies use two different approaches: untargeted and targeted. The untargeted approach involves comprehensive analysis without prior knowledge of the metabolites present in the sample. It encompasses a qualitative or semi-quantitative search of altered metabolites following an external intervention as compared to a control or untreated sample, and it may focus on intracellular (fingerprinting) or extracellular (footprinting) metabolites. The targeted approach focuses on the quantitative analysis of selected metabolites with similar chemical properties (Klassen et al. 2017; Raterink et al. 2014; Villas-Boas et al. 2005). A general metabolomics workflow (Fig. 13.1) encompasses sample collection and pre-treatment, interruption of enzymatic activity (metabolic quenching), and sample extraction. Selective or nonselective extraction procedures, cell disruption, and protein precipitation can also be applied during sample preparation and extraction steps (Álvarez-Sánchez et al. 2010; Raterink et al. 2014).

Data processing and statistical analysis require sophisticated software and chemometric tools to extract the data and to find a discriminant group of metabolites, responsible for the overall alterations in the studied system. The correlation of the results with biochemical routes is the last and probably most essential step in the metabolomics workflow since it can address biological questions (Klassen et al. 2017).

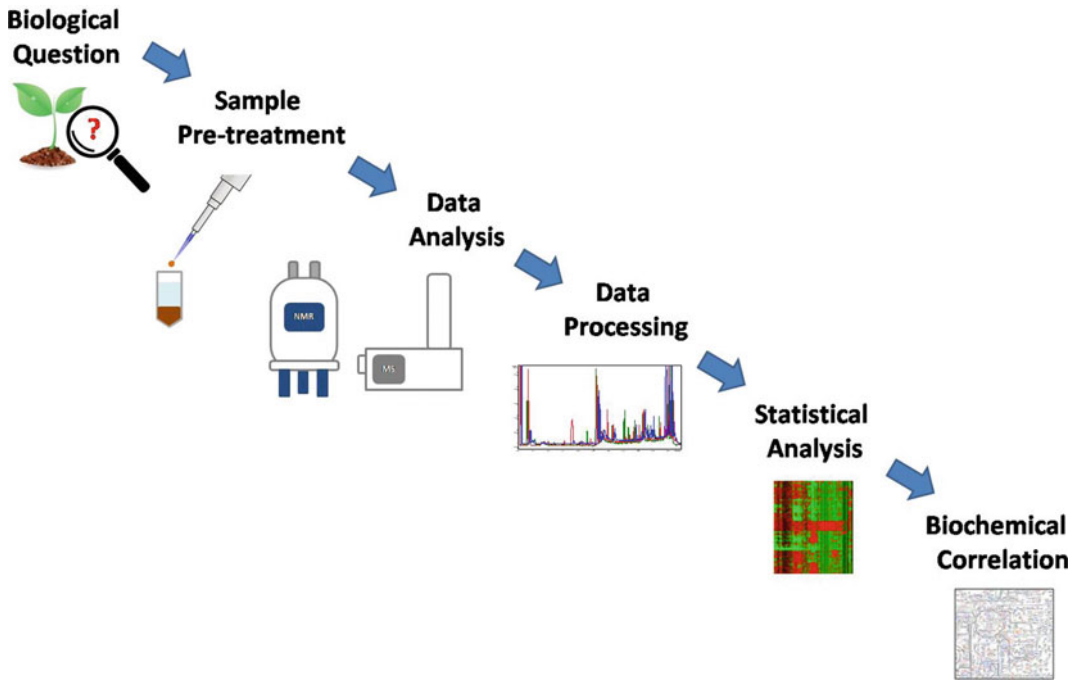

Fig. 13.1 Schematic representation of a general metabolomics workflow

A major challenge in metabolomics studies is to cover the largest possible portion of the metabolome of a biological system, and there is no single analytical technique able to do so. Therefore, it is worthwhile to use more than one technique to obtain a more complete picture of the metabolome of the studied biological system. Nuclear magnetic resonance (NMR) and mass spectrometry (MS) are the most used techniques for metabolite detection (Kuehnbaum and Britz-McKibbin 2013) because they provide complementary information regarding the chemical characteristics of the detected metabolites. Therefore, NMR and MS are frequently applied to plant metabolomics studies (Ernst et al. 2014).

The use of NMR in metabolomics has grown in recent years due to its robustness for structural characterization of unknown metabolites. Other advantages of NMR are its amenability to handle small samples and its nondestructiveness (Putri et al. 2013). However, to obtain a detailed metabolic characterization, it is necessary to use high-resolution equipment, which makes this an expensive technique (Psychogios et al. 2011; Villas-Boas et al. 2005).

MS has some advantages relative to NMR such as its higher sensitivity and specificity, as well as a higher efficiency (Klassen et al. 2017). The combination of MS with separation techniques, such as gas and liquid chromatography (GC-MS and LC-MS, respectively), and capillary electrophoresis (CE-MS), increases the range of metabolite detection of MS for diverse applications of metabolomics studies (de Raad et al. 2016).

GC-MS was one of the first separation techniques used for metabolomics. In this approach, metabolites are separated based on their different boiling points and polarity. The use of capillary columns enhances the resolution and efficiency of this separation technique. A significant advantage of the GC-MS is the use of electron ionization (EI) source with constant energy, which generates highly reproducible MS spectra (Garcia and Barbas 2011) and facilitates the identification of metabolites making it possible to create in-house spectrum libraries. On the other hand, GC-MS analysis can only detect volatile and thermally stable compounds, which may require sample derivatization in order to

make them suitable for the GC-MS analysis (Koek et al. 2011).

LC-MS is a versatile tool, and it has been extensively applied in metabolomics, offering robustness, high sensitivity, and selectivity, besides being easy to operate (Xiayan and Legido-Quigley 2008). Currently, it is the analytical technique most used in metabolomics studies due to the availability of numerous types of chromatographic columns with different stationary phases, from nonpolar to polar, which makes this separation technique very versatile, allowing the identification of a wide range of metabolites (Kloos et al. 2013; Theodoridis et al. 2012).

CE-MS is a powerful technique that facilitates the separation and analysis of polar or charged metabolites in the presence of an electric field. A significant advantage of CE-MS is its requirement of small amounts of sample for analysis. On the other hand, its small sample volume impairs the coupling between the CE and MS, requiring the addition of volatile auxiliary liquids, which will maintain the electrical contact between the separation (CE) and the detection (MS) system in order to facilitate the introduction and ionization of the sample in the mass spectrometer. The application of sheath liquids leads to a decrease in sensitivity by dilution of the metabolites, especially for those metabolites present in low concentrations. The sheath liquid usually consists of a 50% methanol solution, acidified with formic acid or a volatile buffer (such as ammonium acetate, for example). This liquid is mixed with the effluent of the CE in the MS ionization source and nitrogen (nebulizer gas) assisting the formation of the spray (Barbas et al. 2011; Garcia et al. 2017).

Metabolomics studies provide vital information for the understanding of biological processes and has been applied in plant studies (Theodoridis et al. 2008), the analysis of environmental and genetic alterations (Sumner et al. 2015), agricultural research (Tian et al. 2016), treatment and diagnosis of diseases (Gomez-Casati et al. 2013), human health (Manach et al. 2009), response to stress (Obata and Fernie 2012), natural products discovery (Kell 2006). Castor bean metabolomics studies have focused on the response of seeds and seedlings to abiotic stresses and mainly temperature and will be discussed in the following sections.

13.1.2 Perception of Heat Stress by Plants

Higher plants are sessile organisms and, therefore, cannot simply escape from their surroundings when unfavorable environmental conditions arise. As an adaptation to changing environmental conditions, plants have developed a series of biochemical and molecular responses to stress (Shao et al. 2007). Tolerance to abiotic stresses is a very complex phenomenon, in part because stress may occur at multiple developmental stages and plants are often simultaneously affected by more than one type of stress (Chinnusamy et al. 2004). The way in which plants perceive abiotic stresses and switch on adaptive biochemical, molecular, physiological, and morphological responses is critical for their survival in natural environments (Chinnusamy et al. 2004). Although high temperatures usually have a negative effect on plant growth and development (Buchwal et al. 2013; Zhang et al. 2013), plants possess a number of adaptive and acclimation mechanisms to cope with heat stress (Černý et al. 2014; Hasanuzzaman et al. 2013; Lipiec et al. 2013; Yamori et al. 2014). Although high temperature stress responses may vary with the degree and duration of exposure to heat, it mainly depends on the genetic factors. As a result, differences in stress tolerance between genotypes or between developmental stages within a single genotype may arise from differences in signal perception and transduction mechanisms (Chinnusamy et al. 2004). In this context, plant survival under unfavorable environmental conditions will depend on the plant's ability to perceive the stress, generate and transmit the signal, and initiate appropriate physiological and biochemical changes (Hasanuzzaman et al. 2013). For this reason, it is difficult to draw a line to define which temperature is considered as a stress condition for plants. For example, castor bean seeds were shown to

germinate faster and at higher rates at 35 °C as compared to 20 and 25 °C. However, faster germination did not necessarily imply a better performance during seedling establishment. In fact, seedling survival was dramatically reduced at elevated temperatures (Ribeiro et al. 2015c). Nevertheless, metabolomics analysis provided important insights into the understanding of plants stress tolerance under unfavorable environmental.

13.1.3 Metabolomics Is an Important Tool to Investigate Abiotic Stress Tolerance in Plants

Metabolite profiling has been used to dissect seed germination and seedling establishment under normal and stressed conditions in several plant species (Howell et al. 2009; Hu et al. 2018; Mibei et al. 2018; Qiu et al. 2018; Ribeiro et al. 2014, 2015b, c, d; Weitbrecht et al. 2011; Zhao et al. 2013, 2014). For example, Fait et al. (2006) studied the metabolite profile of *Arabidopsis thaliana* seeds throughout the development and germination under different temperatures. During seed maturation, a significant reduction in the content of sugars, organic acids, and amino acids was observed in seeds, which suggested their efficient incorporation into storage into protein and carbohydrate reserves. The transition from reserve accumulation to seed desiccation is an important step during seed maturation, and a major metabolic switch occurs, resulting in the accumulation of sugars, organic acids, nitrogen-rich amino acids, and shikimate-derived metabolites. In contrast, seed stratification, which consists of an additional step during seed germination aiming at breaking dormancy, was associated with a decrease in the concentration of several of the metabolic intermediates that had accumulated during seed desiccation, implying that these intermediates might support a metabolic reorganization needed for seed germination. Concomitantly, the levels of other metabolites significantly increased during stratification and boosted up further during germination (Fait et al. 2006).

Metabolite profiling of plants growing under abiotic stress conditions has provided important information regarding biochemical and molecular changes related to adaptation to different environmental conditions such as drought (Figueroa-Pérez et al. 2014; Wenzel et al. 2014), elevated salinity (Canam et al. 2013; Wu et al. 2013; Zhao et al. 2014) and temperature stress (Guy et al. 2008; Krasensky and Jonak 2012; Zhou et al. 2011). Carbohydrate and amino acid metabolism appear to be part of the mechanisms by which plants adapt to changes in temperature (Diamant et al. 2001; Gray and Heath 2005; Kaplan et al. 2004; Panikulangara et al. 2004; Rizhsky et al. 2004). Kaplan et al. (2004) performed a metabolite-profiling study to determine temporal metabolite dynamics associated with the induction of acquired heat or freezing tolerance in response to heat or cold shock, respectively. Curiously, cold shock had a more significant influence on metabolism than heat shock: The steady state of 311 and 143 metabolites varied in response to cold and heat shock, respectively (Kaplan et al. 2004). A coordinated increase occurred in the pool sizes of pyruvate- and oxaloacetate-derived amino acids, polyamine precursors, and compatible solutes during both heat and cold shocks. Gray and Heath (2005) examined the effects of cold acclimation on the *Arabidopsis* metabolome using a nontargeted metabolic fingerprinting approach. Cold acclimation turned out to involve many complex biochemical changes in the metabolome (Gray and Heath 2005).

Although metabolomics has been used to dissect plant responses to abiotic stresses, most studies of the temperature effect on seedling performance have focused on the ability of plants to maintain homeostasis at extremely cold (0–15 °C) or extremely high (35–50 °C) temperature stress (Guy et al. 2008; Obata and Fernie 2012). Plant metabolic plasticity in response to mildly elevated temperatures (20–35 °C) has received much less attention although it is an essential trait for crop species (Lewicka and Pietruszka 2006).

13.1.4 Castor Bean: An Oilseed Crop with Great Agricultural and Biotechnological Potential

Castor bean is a member of the *Euphorbiaceae* and an important oilseed crop. This species is predominantly cultivated throughout the tropical and subtropical regions of the world. Despite what its common name suggests, its seeds are not real beans and they are not edible. In fact, castor bean seeds contain high levels of ricin, ricinine, and certain allergens that are highly toxic to humans and animals (Ogunniyi 2006; Severino et al. 2012). Pharmaceutical companies and industries use the oil extracted from castor bean seeds due to its unique chemical composition (Ogunniyi 2006; Severino et al. 2012).

Castor bean can grow in dry and hot environments where other crops would not grow and still produce good yield (Vijaya Kumar et al. 1997). This species is therefore considered tolerant to a diverse range of environmental stresses, including drought, heat, and salt (Silva César and Otávio Batalha 2010). For example, castor bean plants are able to partially maintain their photosynthetic functions when subjected to periods of severe drought, but as soon as the drought stress is relieved, they can fully recover their photosynthetic machinery within one day (Sausen and Rosa 2010). Tolerance mechanisms such as early and efficient stomatal control are used by castor bean plants in order to minimize water loss by transpiration and demand for soil water (Sausen and Rosa 2010). Furthermore, castor bean can stand soil contamination by heavy metals such as lead, nickel, and zinc (Liu et al. 2008; Romeiro et al. 2006). Since castor bean possesses the ability to grow in adverse environments, it makes this species an ideal candidate to provide a better understanding of seed germination, seedling performance, and adaptation to high temperature. Small farmers worldwide could benefit from a better understanding of the biochemical and physiological aspects of germination and seedling growth, which is crucial for breeding of high-yielding varieties adapted to various growing environments.

13.1.5 Metabolomics as an Important Tool for Castor Bean Breeding

Genetic variation is fundamental not only to allow plants to adapt to their environment, but also for crop improvement through breeding. Therefore, studying castor bean natural diversity has been the first step toward efficient breeding programs. Even though castor bean is an important nonedible oilseed crop whose oil confers unique properties to fine chemicals and biodiesel produced from it, it has hardly been the subject of molecular and genetic studies. Genotyping of different castor bean germplasm performed and Allan et al. (2008) and Foster et al. (2010) have shown limited genetic diversity, most likely due to the intense selection of cultivars during domestication. The castor bean genome was the first to be sequenced within the *Euphorbiaceae* family, which also includes important crops like rubber tree and cassava (Chan et al. 2010). The availability of a genome sequence and molecular information on genetic variation can help in the improvement of castor bean productivity, but additional efforts are needed to achieve its full potential. For example, only one genetic map for castor bean has been published (Liu et al. 2016). Probably, this lack of rich genetic resources has slowed down castor bean breeding in recent decades, and no large-scale effective breeding program has been developed so far. Thus, genetic and molecular characterization of castor bean accessions available in germplasm banks and production units will be important to support the development of new elite lines and castor oil production systems (Nass et al. 2007).

Together with other omics technologies like genomics, transcriptomics, and proteomics, metabolomics is an important tool for crop improvement and breeding for higher-yielding castor bean cultivars, particularly for adjusting oil composition (Kumar et al. 2017). These technologies can help augment seed oil content at different levels. For instance, castor oil production can be improved by increasing the number of seeds per plant, seed size, or oil content. These

traits are heavily affected by biotic and abiotic stresses, and therefore, breeding for enhanced stress tolerance is an important goal. Additionally, reducing seed ricin content is an important target for castor bean improvement (Milani and de Medeiros Nóbrega 2013).

Integration of different omics technologies, especially metabolomics and transcriptomics, constitutes a promising approach to provide new insights to advance toward manipulating metabolic fluxes and increasing seed oil content. However, integrated analysis of metabolomics and transcriptomics data is not trivial, and four main strategies have been defined to address this challenge (Cavill et al. 2016). One method uses a simple Pearson or Spearman correlation-based integration of transcriptome and metabolome data sets to identify genes whose transcript levels correlate with a certain metabolite. The identified genes and metabolite are expected to have a functional association. Simple correlation methods can be powerful for the integration of transcriptome data into co-expression networks (Serin et al. 2016). However, metabolomic–transcriptomic data integration is not as straightforward. Jin et al. (2017) used co-expression networks to study cold stress responses in *Nicotiana tabacum,* and it was also used to study the underlying mechanisms of the sensitivity of several rice cultivars to high night temperature (Glaubitz et al. 2017). The second method is called concatenation-based integration, which produces data tables that merge metabolomics and transcriptomics data that are subsequently analyzed with standard techniques such as random forests or self-organizing maps. A Web-based tool to analyze data tables produced by this method is MetaGeneAlyse (Daub et al. 2003). Multivariate-based integration is more complex than the previous two methods, but it is more powerful to find meaningful links between metabolites and transcripts. Common multivariate techniques are principal component analysis (PCA), partial least squares (PLS), and canonical correlation analysis (CCA). CCA is integrated into the free R package integrOmics (Le Cao et al. 2009). An extension of PLS named O2PLS seems especially useful for the integration of metabolomics and transcriptomics data (Bylesjo et al. 2007). However, making these statistical tools easily usable by nonexperts poses a big challenge. Whereas the previous three methods integrate metabolomics and transcriptomics data without consideration of available functional knowledge of the genes included in the data, the fourth method, pathway-based integration, takes into account such information, especially genes encoding metabolic enzymes and subsequent metabolites (Cavill et al. 2016). This approach uses automatic mapping of metabolites and transcripts to infer metabolic pathways like those that can be found in databases such as KEGG (Kanehisa and Goto 2000). This method can produce correlations between pathways and phenotypes or information about statistically significant differences in metabolic pathways observed between different treatments or cultivars. This approach has been used to study drought response in *Astragalus membranaceus* Bge. var. *mongolicus*, a Chinese herb (Jia et al. 2016) and seed to seedling transition in *A. thaliana* (Silva et al. 2017).

The integration of metabolomics and transcriptomics data in combination with natural variation of important traits will substantially increase our understanding the molecular mechanisms underlying these traits in castor bean, ultimately advancing toward improving castor oil production while mitigating the problem of ricin.

13.2 Examples of Applications of Metabolomics in Castor Bean Research

13.2.1 Metabolite-Profiling Studies Suggest that Gamma-Aminobutyric Acid (GABA) Acts as a Drought Stress Response Molecule

The first step in the process of seed germination occurs when dry seeds encounter water under favorable conditions. This step is called "seed imbibition," and it is driven by the matrix

potential of the seed. In dry seeds, the matrix potential is very low (down to more than −100 MPa) causing a rapid influx of water (Woodstock 1988). As seed imbibition progresses, the matrix potential increases and the water influx is reduced (Bewley et al. 2013). Measurements of water content in castor bean seeds during imbibition showed that a progressive increase in temperature leads to faster imbibition (Ribeiro 2015), which had two important consequences: It accelerated germination and reduced GABA levels in 6 h-imbibed seeds (Ribeiro 2015). In this section, we discuss the possible role of GABA accumulation during early imbibition and suggest possible implications of faster imbibition on seed germination and seedling establishment.

Several metabolite-profiling studies have shown that GABA accumulates rapidly in plant tissues exposed to a variety of stresses (Bartyzel et al. 2004; Kinnersley and Turano 2000; Locy et al. 1996). Possible functions of GABA include regulation of cytosolic pH, nitrogen storage, plant development, plant defense, and carbon–nitrogen metabolism (Akçay et al. 2012; Bouché and Fromm 2004; Kinnersley and Turano 2000). GABA also acts as an osmolyte to mitigate drought stress (Kramer et al. 2010; Krishnan et al. 2013; Shelp et al. 1999; Vijayakumari and Puthur 2015). Additionally, the GABA shunt is a way to assimilate carbon from glutamate and to generate carbon–nitrogen fluxes that enter the tricarboxylic acid cycle (Bouché and Fromm 2004).

GABA levels were measured by gas chromatography coupled to a quadrupole time of flight mass spectrometry system (GC-TOF-MS) as trimethylsilyl (TMS) ethers derivatives in castor bean seeds that germinated at 20, 25, and 35 °C, as well as during imbibition after they had taken up water to 24.4% of their dry weight (Ribeiro et al. 2015c). Unexpectedly, GABA levels during early castor bean seed imbibition were higher in seeds imbibed at 20 and 25 °C than in seeds imbibed at 35 °C (Ribeiro 2015). To understand these results, it is important to keep in mind that because of the very negative matrix potential of hydrating seeds, initially there is little or no water left for uptake by the embryo cells (Fig. 13.2). Based on their water content, 6 h-imbibed seeds at 35 °C are more hydrated than 6 h imbibed seeds at 20 and 25 °C. Nevertheless, they are not entirely hydrated in either condition. The practical consequence of this differential hydration state is that seeds imbibed at 20 and 25 °C experience a lower matrix potential than seeds imbibed at 35 °C. This reduced hydration state at low temperatures seems to be perceived by the seeds as a drought stress-like condition, which could explain the observed higher GABA accumulation in seeds imbibed at 20 and 25 °C. Several reports have shown that GABA accumulates in response to drought stress (Bor et al. 2009; Krasensky and Jonak 2012; Krishnan et al. 2013; Vijayakumari and Puthur 2015; Wenzel et al. 2014), but none has shown a clear correlation between GABA levels and seed water content.

After radicle protrusion, GABA levels increased in seedlings imbibed at 35 °C as compared with seeds imbibed at 20 and 25 °C (Ribeiro 2015; Ribeiro et al. 2015c). This early temperature stress response was also observed in later stages of seedling development since 14-day-old seedlings grown at 35 °C had higher levels of GABA than 14-day-old seedlings grown at 20 °C. This increase in GABA levels in response to the high temperature correlated with higher expression levels of glutamate decarboxylase gene. Glutamate decarboxylase (GAD) catalyzes the formation of GABA from glutamate and is regulated by the cytosolic concentration of H^+ or Ca^{2+} (Baum et al. 1993, 1996). Increased levels of cytosolic Ca^{2+} lead to the formation of complexes with calmodulin (CaM). Then, the complex Ca^{2+}/CaM activates GAD in the physiological pH range (Baum et al. 1993, 1996). The Ca^{2+}/CaM complex induces the expression of stress-responsive genes and may also be involved in the acquisition of minerals that activate enzymes in stress-related metabolic pathways (Kinnersley and Turano 2000). Therefore, during early seedling development, GABA accumulation seems to occur as a result of heat stress (Ribeiro 2015).

GABA may protect biological membranes and balance the decrease in water potential that

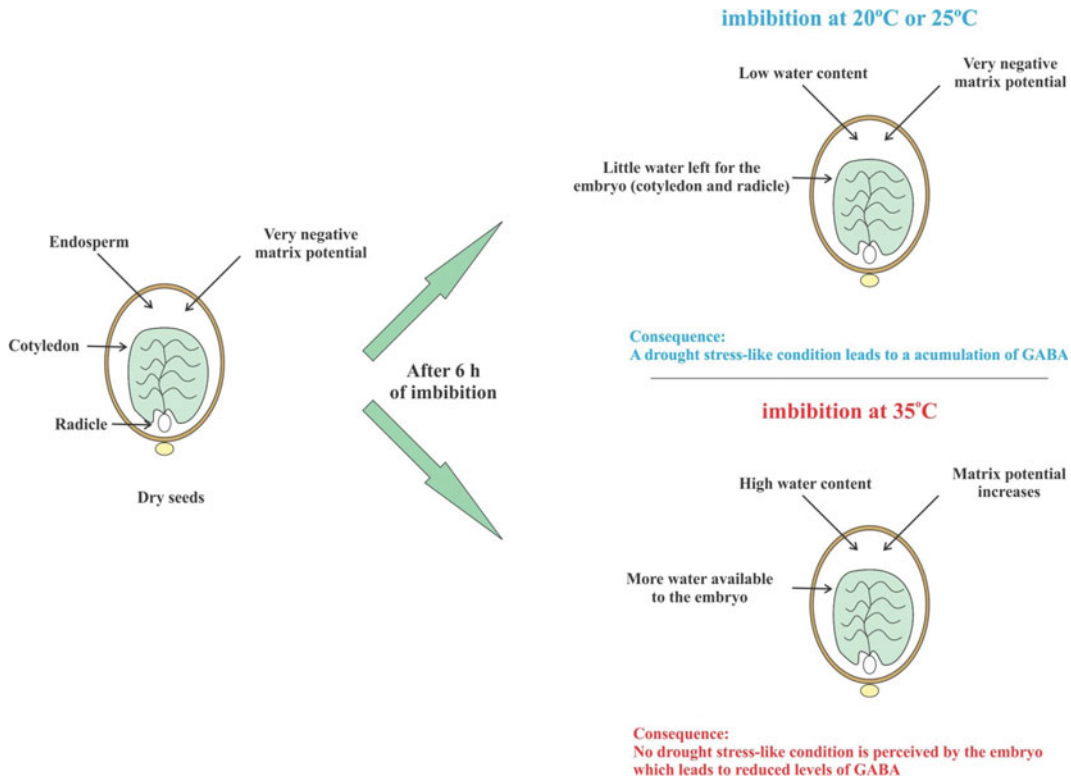

Fig. 13.2 Schematic representation of changes in water content of castor bean seeds during early imbibition and relationship with the matrix potential and GABA accumulation in response to temperature

occurs upon drought stress (Heber et al. 1971; Kinnersley and Turano 2000). The mechanism by which GABA accumulation could mitigate plant stress relies on the linkage between stress perception and physiological responses.

13.2.2 Metabolic Signatures of Castor Bean Seeds and Seedlings in Response to Temperature

Mobilization of seed storage compounds during germination is a crucial process to set proper conditions for seedling establishment (Graham 2008; Weitbrecht et al. 2011). Initial imbibition is often accompanied by large metabolic changes, followed by an increase in respiratory metabolism and gene expression (Bewley et al. 2013; Botha et al. 1992; Sánchez-Linares et al. 2012; Vicente-Carbajosa and Carbonero 2005). For example, metabolite profiling of rice embryos during germination showed that 1 h after imbibition rapid changes in metabolism occurred, including increase in hexose phosphates, tricarboxylic acid cycle intermediates, and GABA (Howell et al. 2009).

Germination in oilseeds is accompanied by the mobilization of storage lipids from the oil bodies located in the endosperm (Yaniv et al. 1998). Triacylglycerol (TAG) is a major seed storage reserve in oilseed species, which accumulates during seed development. TAG is stored in the seed until germination, after which it is used to fuel initial seedling growth (Graham 2008). TAGs are localized near glyoxysomes, the organelles that contain most of the biochemical machinery required to convert fatty acids from TAG to four-carbon compounds. These smaller

compounds are then converted to soluble carbohydrates necessary for seedling growth (Graham 2008). In castor bean, fatty acids are converted to carbohydrates through gluconeogenesis (Kobr and Beevers 1971).

Initial vegetative growth is a crucial stage in the establishment of plants, and little is known about biochemical and molecular changes related to temperature adaptation in castor bean during that stage. Therefore, we applied a multiomics approach involving the use of GC-TOF-MS, GC-MS, and HPLC for metabolite profiling to identify metabolic signatures in response to temperature in such a highly adaptable species. (Ribeiro 2015; Ribeiro et al. 2014, 2015a, b, c, d).

Metabolic pathways of certain carbohydrates and amino acids appear to be a key responsive elements of plasticity and tolerance mechanisms, as suggested by several temperature-responsive metabolite-profiling studies to date (Cook et al. 2004; Guy et al. 2008; Kaplan et al. 2004; Obata and Fernie 2012; Palma et al. 2014; Qi et al. 2011; Rizhsky et al. 2004; Zhang et al. 2013). For example, Rizhsky et al. (2004) assessed the metabolic and molecular responses of *Arabidopsis* to a combination of drought and heat stress, and accumulation of sucrose and other carbohydrates such as glucose, maltose, melibiose, gulose, and mannitol was observed in the treated plants (Rizhsky et al. 2004). A general accumulation of most amino acids and a dramatic decrease in most organic acids and carbohydrates were reported at early time points in rice plants subjected to low temperatures (Zhao et al. 2013). Many studies reported that an indistinct and unidirectional increase in the content of most of the identified metabolites seems to be the main response to a variety of environmental stimuli. For example, both heat stress and cold acclimation in *Arabidopsis* led to a coordinated increase in amino acids, TCA intermediates (fumarate and malate), amine-containing metabolites (β-alanine, GABA, and putrescine), and some carbohydrates, such as maltose, sucrose, raffinose, galactinol, and myo-inositol (Gray and Heath 2005; Kaplan et al. 2004, 2007; Rizhsky et al. 2004).

Carbohydrates, amino acids, and organic acids have been measured in castor bean seeds and seedlings by GC-TOF-MS as TMS derivatives, as well as by high-performance liquid chromatography (HPLC) coupled to an ED40-pulsed electrochemical detector (Dionex). Fatty acids and some secondary metabolites were quantified in castor bean seeds and seedlings by GC-MS (Ribeiro et al. 2014, 2015a, c, d).

Castor bean displayed high plasticity during initial vegetative growth, which was reflected in its seed and seedling metabolomes. An increase in temperature did not lead to an indiscriminate accumulation of the identified metabolites. Instead, castor bean seedlings show a specific metabolic signature to adjust growth and developmental processes in response to higher temperatures, possibly to maintain cellular homeostasis. A shift in carbon–nitrogen metabolism toward the accumulation of carbohydrates at low temperatures and the accumulation of amino acids at high temperatures is an important biochemical response of castor bean seeds and seedlings to temperature. Levels of starch and soluble carbohydrates were up to 20-fold in seeds and seedlings grown at 20 °C than at 35 °C, whereas the levels of amino acids such as methionine, tyrosine, and tryptophan were up to 220-fold under the same conditions. This biochemical response was observed during seed imbibition, germination, as well as early seedling growth (Ribeiro 2015; Ribeiro et al. 2014, 2015b, c, d).

Higher carbohydrate levels in castor bean leaves, roots, and seeds, grown at low temperatures could have resulted from up-regulation of biosynthetic pathways, from down-regulation of catabolic pathways. Some genes involved in starch catabolism were up-regulated in leaves of seedlings grown at 20 °C compared with 35 °C. However, up-regulation of genes involved in starch biosynthesis seems to compensate for this and, therefore, is the likely explanation for higher levels of starch in leaves of seedlings grown at 20 °C (Fig. 13.3). Higher levels of fructose, glucose, and sucrose in leaves of seedlings

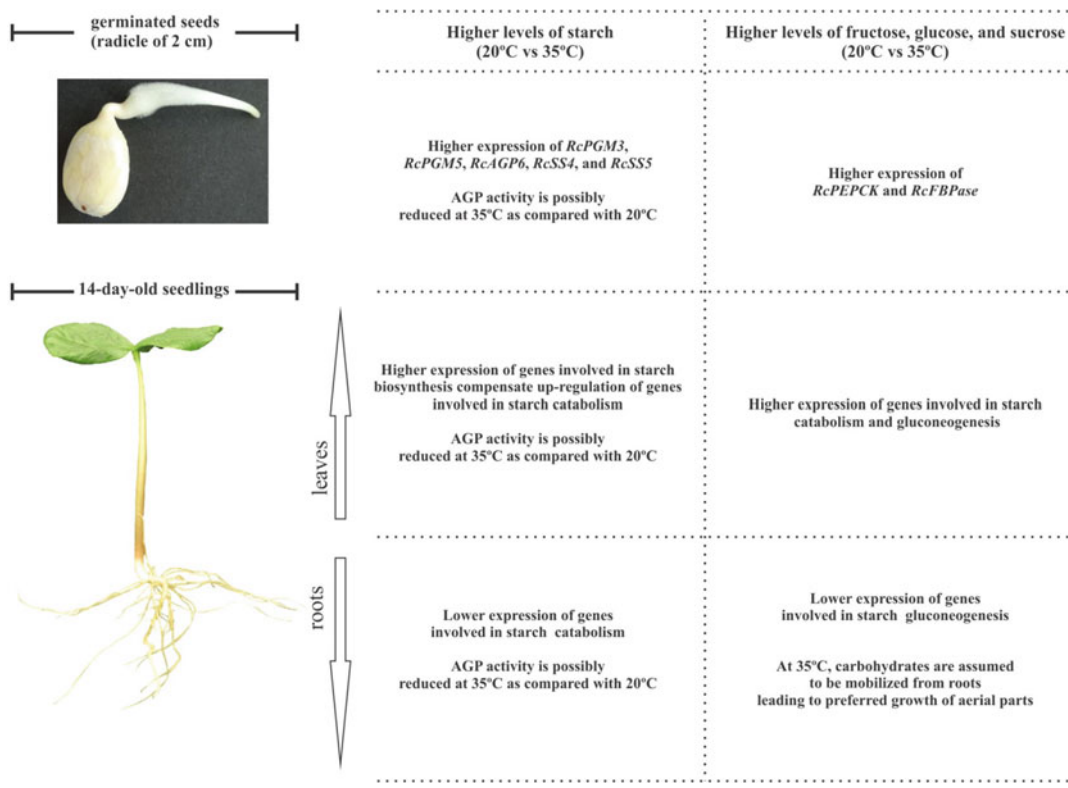

Fig. 13.3 Schematic representation of expression levels of genes encoding key enzymes involved in starch biosynthesis, starch catabolism, and gluconeogenesis and its relation to carbohydrate accumulation at lower temperatures

grown at 20 °C compared with 35 °C may result from a coordinated up-regulation of starch catabolism and gluconeogenesis pathways. Carbohydrates were assumed to be mobilized from the roots to the aerial parts through the hypocotyl, leading to a preferred growth of true leaves, at the expense of the roots, since starch degradation and gluconeogenesis seem to be enhanced in the roots of seedlings grown at 35 °C compared with 20 °C (Fig. 13.3). In seeds, temperature had a greater effect on genes related to starch biosynthesis than on those involved in starch catabolism (Fig. 13.3). Higher expression levels of phosphoglucomutase (*RcPGM*) and starch synthase (*RcSS*) genes is a likely explanation for the higher starch levels observed in seeds germinated at 20 °C, whereas higher expression levels of phosphoenolpyruvate carboxykinase (*RcPEPCK*) and fructose-1,6-bisphosphatase (*RcFBPase*) genes may explain the higher concentrations of fructose, glucose, and sucrose (Ribeiro 2015; Ribeiro et al. 2015a). Although the physiological role of these metabolic changes is not fully understood, these results provide valuable insights for the understanding of seed germination and seedling establishment in response to temperature.

13.2.3 Other Metabolomics Studies on Castor Bean

Castor bean seeds contain high amounts of ricin, a ribosome-inactivating protein (RIP) constituted by two protein chains. The A chain is a ribosome-inactivating enzyme, which inhibits protein synthesis in eukaryotic cells leading to apoptosis, whereas the B chain is responsible for binding to cell surface receptors and delivering the active A chain into the cell (Guo et al. 2014).

The Chemical Weapons Convention considers this protein a powerful biological agent used for bioterrorism (Ovenden et al. 2010). For this reason, the development of analytical methods to determine which cultivar the toxin was isolated from would be significantly useful to law enforcement and forensic agencies (Ovenden et al. 2014). ^{1}H NMR spectroscopy was used to differentiate and characterize eight provenances of six castor bean cultivars based on extracts produced from whole seeds. Metabolite-profiling analysis allowed the identification of ricinin, *N*- and *O*-demethylricinin, sucrose, and phenylalanine as important molecular markers for characterization of cultivars. Orthogonal projections to latent structures discriminant analysis (OPLS-DA) allowed successful separation of the specimens according to their provenance (Pigott et al. 2012). Metabolomics studies also differentiated castor bean provenance and cultivars for forensic applications. For that purpose, the authors used HPLC with a UV detector (HPLC-UV), liquid chromatography–mass spectrometry (LC-MS) and ^{1}H NMR spectroscopy. Interestingly, they were able to dissect the chemical composition to the level of specimen differentiation (Ovenden et al. 2010). In addition, multiple reaction monitoring mass spectrometry (MRM–MS) was used to quantify ricin and castor bean agglutinin (RCA) content in 18 cultivars, since this method allowed to distinguish both ricin A-, ricin B-, and RCA-specific peptides (Schieltz et al. 2015). Laser ablation-ICP-MS was used to perform trace isotope elemental fingerprinting of castor bean seeds for provenance determination (Bagas et al. 2017). Out of 92 isotopes measured, fifteen (^{24}Mg, ^{27}Al, ^{44}Ca, ^{53}Cr, ^{55}Mn, ^{57}Fe, ^{60}Ni, ^{65}Cu, ^{66}Zn, ^{75}As, ^{85}Rb, ^{88}Sr, ^{98}Mo, ^{138}Ba, and ^{202}Hg) yielded data relevant to all collection sites. Those data were used to establish the provenance of seeds of "unknown" origin. The data were further analyzed using multivariate statistical analysis, which allowed the identification of unique provenance isotopes profiles (Bagas et al. 2017).

^{1}H NMR spectroscopy analysis of biofluid and tissues provides a rapid, nondestructive, and high-throughput method for evaluation of chronic toxicity. In a pioneer study, the toxicity of a ricin-enriched extract (REE) obtained from castor bean seeds was investigated using an NMR-based metabolomics approach complemented with histopathological inspection and clinical chemistry. The authors reported that chronic administration of REE could cause kidney and lung impairment, spleen and thymus dysfunction, and diminished nutrient intake in rats. This metabolomic approach provided a systematic and holistic view of this model organism's response to drugs. The authors also claimed that this approach was suitable for dynamic toxicological studies (Guo et al. 2014). In a subsequent study, the same authors orally administered a crude ricinine extract to rats and analyzed their urine, serum, and kidney samples by ^{1}H NMR metabolomics (Guo et al. 2015). They also performed histopathological inspection and biochemical assays to demonstrate that administration of crude ricinine produced obvious nephrotoxicity and severe metabolic alterations (Guo et al. 2015).

Liquid chromatography–tandem mass spectrometry (LC-MS/MS) analysis of four castor bean hybrids detected 60 metabolites of high commercial value, associated to primary and secondary metabolism, including amino acids, fatty acids, flavonols, flavones, flavanones, phenyl propanoids, and dopamine. Furthermore, it was found that dopamine and tyrosine gave rise to hydroxytyrosol in the seeds of some hybrids (Merkouropoulos et al. 2016).

13.3 Conclusions

In general, metabolomics has the advantage of the relatively smaller number of relevant metabolites, which makes it easier to extract information to answer biological questions. Additionally, metabolites are the downstream products of cellular activity and represent the immediate cellular response to external stimuli. The use of metabolomics is becoming increasingly popular toward understanding of plant plasticity under stress conditions due to its unique ability to generate functional and systems

data. The castor bean metabolome has been vastly investigated, especially with regard to its ability to respond to external stimuli. These results will help us understand the molecular requirements for vigorous castor bean seed germination and seedling growth under different environmental conditions.

References

Akçay N, Bor M, Karabudak T, Özdemir F, Türkan T (2012) Contribution of Gamma amino butyric acid (GABA) to salt stress responses of *Nicotiana sylvestris* CMSII mutant and wild type plants. J Plant Physiol 169(5):452–458

Allan G, Williams A, Rabinowicz PD, Chan AP, Ravel J, Keim P (2008) Worldwide genotyping of castor bean germplasm (*Ricinus communis* L.) using AFLPs and SSRs. Genet Resour Crop Evol 55(3):365–378

Álvarez-Sánchez B, Priego-Capote F, Luque de Castro MD (2010) Metabolomics analysis I. Selection of biological samples and practical aspects preceding sample preparation. Trends Analyt Chem 29(2):111–119

Bagas CK, Scadding RL, Scadding CJ, Watling RJ, Roberts W, Ovenden SPB (2017) Trace isotope analysis of *Ricinus communis* seed core for provenance determination by laser ablation-ICP-MS. Forensic Sci Int 270:46–54

Barbas C, Moraes EP, Villasenor A (2011) Capillary electrophoresis as a metabolomics tool for non-targeted fingerprinting of biological samples. J Pharmaceut Biomed Anal 55(4):823–831

Bartyzel I, Pelczar K, Paszkowski A (2004) Functioning of the γ-aminobutyrate pathway in wheat seedlings affected by osmotic stress. Biol Plant 47(2):221–225

Baum G, Chen Y, Arazi T, Takatsuji H, Fromm H (1993) A plant glutamate decarboxylase containing a calmodulin binding domain: Cloning, sequence, and functional analysis. J Biol Chem 268(26):19610–19617

Baum G, Lev-Yadun S, Fridmann Y, Arazi T, Katsnelson H, Zik M, Fromm H (1996) Calmodulin binding to glutamate decarboxylase is required for regulation of glutamate and GABA metabolism and normal development in plants. EMBO J 15(12):2988–2996

Bewley JD, Bradford KJ, Hilhorst HWM, Nonogaki H (2013) Seeds—physiology of development, germination and dormancy. Springer, New York

Bor M, Seckin B, Ozgur R, Yilmaz O, Ozdemir F, Turkan I (2009) Comparative effects of drought, salt, heavy metal and heat stresses on gamma-aminobutryric acid levels of sesame (*Sesamum indicum* L.). Acta Physiol Plant 31(3):655–659

Botha FC, Potgieter GP, Botha AM (1992) Respiratory metabolism and gene expression during seed germination. Plant Growth Regul 11(3):211–224

Bouché N, Fromm H (2004) GABA in plants: Just a metabolite? Trends Plant Sci 9(3):110–115

Buchwal A, Rachlewicz G, Fonti P, Cherubini P, Gärtner H (2013) Temperature modulates intra-plant growth of Salix polaris from a high Arctic site (Svalbard). Polar Biol 36(9):1305–1318

Bylesjo M, Eriksson D, Kusano M, Moritz T, Trygg J (2007) Data integration in plant biology: the O2PLS method for combined modeling of transcript and metabolite data. Plant J 52(6):1181–1191

Canam T, Li X, Holowachuk J, Yu M, Xia J, Mandal R, Krishnamurthy R, Bouatra S, Sinelnikov I, Yu B, Grenkow L, Wishart DS, Steppuhn H, Falk KC, Dumonceaux TJ, Gruber MY (2013) Differential metabolite profiles and salinity tolerance between two genetically related brown-seeded and yellow-seeded *Brassica carinata* lines. Plant Sci 198:17–26

Canuto GAB, da Cruz PLR, Faccio AT, Klassen A, Tavares MFM (2015) Neglected diseases prioritized in Brazil under the perspective of metabolomics: a review. Electrophoresis 36(18):2336–2347

Cavill R, Jennen D, Kleinjans J, Briedé JJ (2016) Transcriptomic and metabolomic data integration. Brief Bioinformat 17(5):891–901

Černý M, Jedelský PL, Novák J, Schlosser A, Brzobohatý B (2014) Cytokinin modulates proteomic, transcriptomic and growth responses to temperature shocks in *Arabidopsis*. Plant Cell Environ 37(7):1641–1655

Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, Melake-Berhan A, Jones KM, Redman J, Chen G, Cahoon EB, Gedil M, Stanke M, Haas BJ, Wortman JR, Fraser-Liggett CM, Ravel J, Rabinowicz PD (2010) Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotechnol 28:951

Chinnusamy V, Schumaker K, Zhu JK (2004) Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. J Exp Bot 55 (395):225–236

Cook D, Fowler S, Fiehn O, Thomashow MF (2004) A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of Arabidopsis. Proc Natl Acad Sci USA 101(42):15243–15248

Daub CO, Kloska S, Selbig J (2003) MetaGeneAlyse: analysis of integrated transcriptional and metabolite data. Bioinformatics 19(17):2332–2333

de Raad M, Fischer CR, Northen TR (2016) High-throughput platforms for metabolomics. Curr Opin Chem Biol 30:7–13

Diamant S, Eliahu N, Rosenthal D, Goloubinoff P (2001) Chemical chaperones regulate molecular chaperones in vitro and in cells under combined salt and heat stresses. J Biol Chem 276(43):39586–39591

Ernst M, Silva DB, Silva RR, Vencio RZ, Lopes NP (2014) Mass spectrometry in plant metabolomics strategies: from analytical platforms to data acquisition and processing. Nat Prod Rep 31(6):784–806

Fait A, Angelovici R, Less H, Ohad I, Urbanczyk-Wochniak E, Fernie AR, Galili G (2006)

Arabidopsis seed development and germination is associated with temporally distinct metabolic switches. Plant Physiol 142(3):839–854

Fiehn O (2001) Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. Comp Funct Genom 2(3):155–168

Figueroa-Pérez MG, Rocha-Guzmán NE, Pérez-Ramírez IF, Mercado-Silva E, Reynoso-Camacho R (2014) Metabolite profile, antioxidant capacity, and inhibition of digestive enzymes in infusions of peppermint (*Mentha piperita*) grown under drought stress. J Agric Food Chem 62(49):12027–12033

Foster JT, Allan GJ, Chan AP, Rabinowicz PD, Ravel J, Jackson PJ, Keim P (2010) Single nucleotide polymorphisms for assessing genetic diversity in castor bean (*Ricinus communis*). BMC Plant Biol 10(1):13

Garcia A, Barbas C (2011) Gas chromatography-mass spectrometry (GC-MS)-based metabolomics. Meth Mol Biol 708:191–204

Garcia A, Godzien J, Lopez-Gonzalvez A, Barbas C (2017) Capillary electrophoresis mass spectrometry as a tool for untargeted metabolomics. Bioanalysis 9 (1):99–130

Glaubitz U, Li X, Schaedel S, Erban A, Sulpice R, Kopka J, Hincha DK, Zuther E (2017) Integrated analysis of rice transcriptomic and metabolomic responses to elevated night temperatures identifies sensitivity- and tolerance-related profiles. Plant Cell Environ 40(1):121–137

Gomez-Casati DF, Zanor MI, Busi MV (2013) Metabolomics in plants and humans: applications in the prevention and diagnosis of diseases. Biomed Res Int 2013:792527

Graham IA (2008) Seed storage oil mobilization. Annu Rev Plant Biol 59:115–142

Gray GR, Heath D (2005) A global reorganization of the metabolome in *Arabidopsis* during cold acclimation is revealed by metabolic fingerprinting. Physiol Plant 124(2):236–248

Guo P, Wang J, Dong G, Wei D, Li M, Yang M, Kong L (2014) NMR-based metabolomics approach to study the chronic toxicity of crude ricin from castor bean kernels on rats. Mol BioSyst 10(9):2426–2440

Guo P, Wei D, Wang J, Dong G, Zhang Q, Yang M, Kong L (2015) Chronic toxicity of crude ricinine in rats assessed by ^{1}H NMR metabolomics analysis. RSC Adv 5(34):27018–27028

Guy C, Kaplan F, Kopka J, Selbig J, Hincha DK (2008) Metabolomics of temperature stress. Physiol Plant 132 (2):220–235

Hasanuzzaman M, Nahar K, Alam MM, Roychowdhury R, Fujita M (2013) Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants. Intl J Mol Sci 14(5):9643–9684

Heber U, Tyankova L, Santarius KA (1971) Stabilization and inactivation of biological membranes during freezing in the presence of amino acids. BBA Biomembr 241(2):578–592

Howell KA, Narsai R, Carroll A, Ivanova A, Lohse M, Usadel B, Millar AH, Whelan J (2009) Mapping metabolic and transcript temporal switches during germination in rice highlights specific transcription factors and the role of RNA instability in the germination process. Plant Physiol 149(2):961–980

Hu R, Zhang Y, Chen Y, Lin G (2018) Dynamic metabolic profiling in vegetable soybean seed development. Emirates J Food Agri 30(2):90–98

Jia X, Sun C, Zuo Y, Li G, Li G, Ren L, Chen G (2016) Integrating transcriptomics and metabolomics to characterise the response of *Astragalus membranaceus* Bge. var. mongolicus (Bge.) to progressive drought stress. BMC Genomics 17:188

Jin J, Zhang H, Zhang J, Liu P, Chen X, Li Z, Xu Y, Lu P, Cao P (2017) Integrated transcriptomics and metabolomics analysis to characterize cold stress responses in Nicotiana tabacum. BMC Genomics 18:496

Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucl Acids Res 28(1):27–30

Kaplan F, Kopka J, Haskell DW, Zhao W, Schiller KC, Gatzke N, Sung DY, Guy CL (2004) Exploring the temperature-stress metabolome of *Arabidopsis*. Plant Physiol 136(4):4159–4168

Kaplan F, Kopka J, Sung DY, Zhao W, Popp M, Porat R, Guy CL (2007) Transcript and metabolite profiling during cold acclimation of *Arabidopsis* reveals an intricate relationship of cold-regulated gene expression with modifications in metabolite content. Plant J 50(6):967–981

Kell DB (2006) Systems biology, metabolic modelling and metabolomics in drug discovery and development. Drug Discov Today 11(23):1085–1092

Kinnersley AM, Turano FJ (2000) Gamma aminobutyric acid (GABA) and plant responses to stress. Crit Rev Plant Sci 19(6):479–509

Klassen A, Faccio AT, Canuto GAB, da Cruz PLR, Ribeiro HC, Tavares MFM, Sussulini A (2017) Metabolomics: definitions and significance in systems biology. Metabolomics: from fundamentals to clinical applications. Springer International Publishing, Cham, pp 3–17

Kloos D-P, Lingeman H, Niessen WMA, Deelder AM, Giera M, Mayboroda OA (2013) Evaluation of different column chemistries for fast urinary metabolic profiling. J Chromatogr B 927:90–96

Kobr MJ, Beevers H (1971) Gluconeogenesis in Castor bean endosperm. 1. Changes in glycolytic intermediates. Plant Physiol 47(1):48–52

Koek MM, Jellema RH, van der Greef J, Tas AC, Hankemeier T (2011) Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. Metabolomics 7(3):307–328

Kramer D, Breitenstein B, Kleinwchter M, Selmar D (2010) Stress metabolism in green coffee beans (*Coffea arabica* L.): expression of dehydrins and accumulation of GABA during drying. Plant Cell Physiol 51(4):546–553

Krasensky J, Jonak C (2012) Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. J Exp Bot 63(4):1593–1608

Krishnan S, Laskowski K, Shukla V, Merewitz EB (2013) Mitigation of drought stress damage by exogenous application of a non-protein amino acid γ-aminobutyric acid on perennial ryegrass. J Amer Soc Hortic Sci 138(5):358–366

Kuehnbaum NL, Britz-McKibbin P (2013) New advances in separation science for metabolomics: resolving chemical diversity in a post-genomic era. Chem Rev 113(4):2437–2468

Kumar R, Bohra A, Pandey AK, Pandey MK, Kumar A (2017) Metabolomics for plant improvement: status and prospects. Front Plant Sci 8:1302

Le Cao KA, Gonzalez I, Dejean S (2009) integrOmics: an R package to unravel relationships between two omics datasets. Bioinformatics 25(21):2855–2856

Lewicka S, Pietruszka M (2006) Theoretical search for the growth-temperature relationship in plants. Gen Physiol Biophys 25(2):125–136

Lipiec J, Doussan C, Nosalewicz A, Kondracka K (2013) Effect of drought and heat stresses on plant growth and yield: A review. Intl Agrophys 27(4):463–477

Liu S, Yin X, Lu J, Liu C, Bi C, Zhu H, Shi Y, Zhang D, Wen D, Zheng J, Cui Y, Li W (2016) The first genetic linkage map of *Ricinus communis* L. based on genome-SSR markers. Indust Crop Prod 89:103–108

Liu X, Gao Y, Khan S, Duan G, Chen A, Ling L, Zhao L, Liu Z, Wu X (2008) Accumulation of Pb, Cu, and Zn in native plants growing on contaminated sites and their potential accumulation capacity in Heqing, Yunnan. J Environ Sci 20(12):1469–1474

Locy RD, Singh NK, Cherry JH (1996) GABA accumulation during heat stress in Arabidopsis. Plant Physiol 111(2):194

Manach C, Hubert J, Llorach R, Scalbert A (2009) The complex links between dietary phytochemicals and human health deciphered by metabolomics. Mol Nutr Food Res 53(10):1303–1315

Merkouropoulos G, Kapazoglou A, Drosou V, Jacobs E, Krolzig A, Papadopoulos C, Hilioti Z (2016) Dwarf hybrids of the bioenergy crop *Ricinus communis* suitable for mechanized harvesting reveal differences in morpho-physiological characteristics and seed metabolic profiles. Euphytica 210(2):207–219

Mibei EK, Owino WO, Ambuko J, Giovannoni JJ, Onyango AN (2018) Metabolomic analyses to evaluate the effect of drought stress on selected African Eggplant accessions. J Sci Food Agric 98(1):205–216

Milani M, de Medeiros Nóbrega MB (2013) Castor breeding. Plant breeding from laboratories to fields. InTech. https://doi.org/10.5772/56216

Nass L, Pereira A, Ellis D (2007) Biofuels in Brazil: an overview. Crop Sci 47(6). https://doi.org/10.2135/cropsci2007.03.0166

Nicholson JK, Lindon JC, Holmes E (1999) 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 29(11):1181–1189

Obata T, Fernie AR (2012) The use of metabolomics to dissect plant responses to abiotic stresses. Cell Mol Life Sci 69(19):3225–3243. https://doi.org/10.1007/s00018-012-1091-5

Ogunniyi DS (2006) Castor oil: a vital industrial raw material. Bioresour Technol 97(9):1086–1091

Ovenden SPB, Gordon BR, Bagas CK, Muir B, Rochfort S, Bourne DJ (2010) A study of the metabolome of *Ricinus communis* for forensic applications. Austral J Chem 63(1):8–21

Ovenden SPB, Pigott EJ, Rochfort S, Bourne DJ (2014) Liquid chromatography-mass spectrometry and chemometric analysis of ricinus communis extracts for cultivar identification. Phytochem Anal 25(5):476–484

Palma F, Carvajal F, Lluch C, Jamilena M, Garrido D (2014) Changes in carbohydrate content in zucchini fruit (*Cucurbita pepo* L.) under low temperature stress. Plant Sci 217–218:78–86

Panikulangara TJ, Eggers-Schumacher G, Wunderlich M, Stransky H, Schöffl F (2004) Galactinol synthase1. A novel heat shock factor target gene responsible for heat-induced synthesis of raffinose family oligosaccharides in arabidopsis. Plant Physiol 136(2):3148–3158

Pigott EJ, Roberts W, Ovenden SPB, Rochfort S, Bourne DJ (2012) Metabolomic investigations of *Ricinus communis* for cultivar and provenance determination. Metabolomics 8(4):634–642

Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, Sinelnikov I, Krishnamurthy R, Eisner R, Gautam B, Young N, Xia J, Knox C, Dong E, Huang P, Hollander Z, Pedersen TL, Smith SR, Bamforth F, Greiner R, McManus B, Newman JW, Goodfriend T, Wishart DS (2011) The human serum metabolome. PLoS ONE 6(2):e16957

Putri SP, Yamamoto S, Tsugawa H, Fukusaki E (2013) Current metabolomics: technological advances. J Biosci Bioeng 116(1):9–16

Qi H, Hua L, Zhao L, Zhou L (2011) Carbohydrate metabolism in tomato (*Lycopersicon esculentum* Mill.) seedlings and yield and fruit quality as affected by low night temperature and subsequent recovery. Afr J Biotechnol 10(30):5743–5749

Qiu Z, Wu X, Zhang J, Huang C (2018) High-temperature induced changes of extracellular metabolites in Pleurotus ostreatus and their positive effects on the growth of *Trichoderma asperellum*. Front Microbiol 9(JAN)

Raterink R-J, Lindenburg PW, Vreeken RJ, Ramautar R, Hankemeier T (2014) Recent developments in sample-pretreatment techniques for mass spectrometry-based metabolomics. TrAC Trends Analyt Chem 61:157–167

Ribeiro PR (2015) Biochemical, physiological and molecular responses of *Ricinus communis* seeds and seedlings to different temperatures: a multi-omics approach. Plant Physiology Department. Wageningen University, The Netherlands, p 203

Ribeiro PR, Fernandez LG, de Castro RD, Ligterink W, Hilhorst HWM (2014) Physiological and biochemical responses of *Ricinus communis* seedlings to different temperatures: a metabolomics approach. BMC Plant Biol 14(1):223

Ribeiro PR, Ligterink W, Hilhorst HWM (2015a) Expression profiles of genes related to carbohydrate metabolism provide new insights into carbohydrate accumulation in seeds and seedlings of Ricinus communis in response to temperature. Plant Physiol Biochem 95:103–112

Ribeiro PR, Willems LAJ, Mudde E, Fernandez LG, de Castro RD, Ligterink W, Hilhorst HWM (2015b) Metabolite profiling of the oilseed crop *Ricinus communis* during early seed imbibition reveals a specific metabolic signature in response to temperature. Indust Crop Prod 67:305–309

Ribeiro PR, Willems LAJ, Mutimawurugo MC, Fernandez LG, de Castro RD, Ligterink W, Hilhorst HWM (2015c) Metabolite profiling of *Ricinus communis* germination at different temperatures provides new insights into thermo-mediated requirements for successful seedling establishment. Plant Sci (under review)

Ribeiro PR, Zanotti RF, Deflers C, Fernandez LG, de Castro RD, Ligterink W, Hilhorst HWM (2015d) Effect of temperature on biomass allocation in seedlings of two contrasting genotypes of the oilseed crop *Ricinus communis*. J Plant Physiol 185:31–39

Rizhsky L, Liang HJ, Shuman J, Shulaev V, Davletova S, Mittler R (2004) When defense pathways collide. The response of *Arabidopsis* to a combination of drought and heat stress. Plant Physiol 134(4):1683–1696

Romeiro S, Lagôa AMMA, Furlani PR, De Abreu CA, De Abreu MF, Erismann NM (2006) Lead uptake and tolerance of *Ricinus communis* L. Brazn J Plant Physiol 18(4):483–489

Sánchez-Linares L, Gavilanes-Ruíz M, Díaz-Pontones D, Guzmán-Chávez F, Calzada-Alejo V, Zurita-Villegas V, Luna-Loaiza V, Moreno-Sánchez R, Bernal-Lugo I, Sánchez-Nieto S (2012) Early carbon mobilization and radicle protrusion in maize germination. J Exp Bot 63(12):4513–4526

Sausen TL, Rosa LMG (2010) Growth and limitations to carbon assimilation in *Ricinus communis* (Euphorbiaceae) under soil water stress conditions. Acta Bot Brasil 24(3):648–654

Schieltz DM, McWilliams LG, Kuklenyik Z, Prezioso SM, Carter AJ, Williamson YM, McGrath SC, Morse SA, Barr JR (2015) Quantification of ricin, RCA and comparison of enzymatic activity in 18 *Ricinus communis* cultivars by isotope dilution mass spectrometry. Toxicon 95:72–83

Serin EAR, Nijveen H, Hilhorst HWM, Ligterink W (2016) Learning from co-expression networks: possibilities and challenges. Front Plant Sci 7:444

Severino LS, Auld DL, Baldanzi M, Cândido MJD, Chen G, Crosby W, Tan D, He X, Lakshmamma P, Lavanya C, Machado OLT, Mielke T, Milani M, Miller TD, Morris JB, Morse SA, Navas AA, Soares DJ, Sofiatti V, Wang ML, Zanotto MD, Zieler H (2012) A review on the challenges for increased production of castor. Agron J 104(4):853–880

Shao HB, Guo QJ, Chu LY, Zhao XN, Su ZL, Hu YC, Cheng JF (2007) Understanding molecular mechanism of higher plant plasticity under abiotic stress. Colloids Surf B 54(1):37–45

Shelp BJ, Bown AW, McLean MD (1999) Metabolism and functions of gamma-aminobutyric acid. Trends Plant Sci 4(11):446–452

Silva AT, Ligterink W, Hilhorst HWM (2017) Metabolite profiling and associated gene expression reveal two metabolic shifts during the seed-to-seedling transition in *Arabidopsis thaliana*. Plant Mol Biol 95(4–5):481–496

Silva César A, Otávio Batalha M (2010) Biodiesel production from castor oil in Brazil: a difficult reality. Energy Policy 38(8):4031–4039

Sumner LW, Lei Z, Nikolau BJ, Saito K (2015) Modern plant metabolomics: advanced natural product gene discoveries, improved technologies, and future prospects. Nat Prod Rep 32(2):212–229

Theodoridis G, Gika HG, Wilson ID (2008) LC-MS-based methodology for global metabolite profiling in metabonomics/metabolomics. TrAC Trends Analyt Chem 27(3):251–260

Theodoridis GA, Gika HG, Want EJ, Wilson ID (2012) Liquid chromatography-mass spectrometry based global metabolite profiling: a review. Analyt Chim Acta 711:7–16

Tian H, Lam SM, Shui G (2016) Metabolomics, a powerful tool for agricultural research. Int J Mol Sci 17(11)

Vicente-Carbajosa J, Carbonero P (2005) Seed maturation: developing an intrusive phase to accomplish a quiescent state. Int J Dev Biol 49(5–6):645–651

Vijaya Kumar P, Ramakrishna YS, Ramana Rao BV, Victor US, Srivastava NN, Subba Rao AVM (1997) Influence of moisture, thermal and photoperiodic regimes on the productivity of castor beans (*Ricinus communis* L.). Agric Forest Meteorol 88(1–4):279–289

Vijayakumari K, Puthur JT (2015) γ-Aminobutyric acid (GABA) priming enhances the osmotic stress tolerance in *Piper nigrum* Linn. plants subjected to PEG-induced stress. Plant Growth Regul

Villas-Boas SG, Mas S, Akesson M, Smedsgaard J, Nielsen J (2005) Mass spectrometry in metabolome analysis. Mass Spectromet Rev 24(5):613–646

Weitbrecht K, Müller K, Leubner-Metzger G (2011) First off the mark: early seed germination. J Exp Botany 62 (10):3289–3309

Wenzel A, Frank T, Reichenberger G, Herz M, Engel KH (2014) Impact of induced drought stress on the metabolite profiles of barley grain. Metabolomics 11 (2):454–467

Woodstock LW (1988) Seed imbibition: a critical period for successful germination. J Seed Technol 12(1):1–15

Wu D, Cai S, Chen M, Ye L, Chen Z, Zhang H, Dai F, Wu F, Zhang G (2013) Tissue metabolic responses to salt stress in wild and cultivated barley. PLoS ONE 8(1)

Xiayan L, Legido-Quigley C (2008) Advances in separation science applied to metabonomics. Electrophoresis 29(18):3724–3736

Yamori W, Hikosaka K, Way DA (2014) Temperature response of photosynthesis in C3, C4, and CAM plants: temperature acclimation and temperature adaptation. Photosynth Res 119(1–2):101–117

Yaniv Z, Shabelsky E, Schafferman D, Granot I, Kipnis T (1998) Oil and fatty acid changes in Sinapis and Crambe seeds during germination and early development. Indust Crop Prod 9(1):1–8

Zhang W, Song L, Teixeira da Silva JA, Sun H (2013) Effects of temperature, plant growth regulators and substrates and changes in carbohydrate content during bulblet formation by twin scale propagation in *Hippeastrum vittatum* 'Red lion'. Sci Hort 160: 230–237

Zhao X, Wang W, Zhang F, Deng J, Li Z, Fu B (2014) Comparative metabolite profiling of two rice genotypes with contrasting salt stress tolerance at the seedling stage. PLoS ONE 9(9)

Zhao XQ, Wang WS, Zhang F, Zhang T, Zhao W, Fu BY, Li ZK (2013) Temporal profiling of primary metabolites under chilling stress and its association with seedling chilling tolerance of rice (*Oryza sativa* L.). Rice 6(1):1–13

Zhou MQ, Shen C, Wu LH, Tang KX, Lin J (2011) CBF-dependent signaling pathway: a key responder to low temperature stress in plants. Crit Rev Biotechnol 31(2):186–192

Biotechnological Means for Genetic Improvement in Castor Bean as a Crop of the Future

14

Mulpuri Sujatha and Muddanuru Tarakeswari

Abstract

Profitable cultivation of castor bean is beset with problems of vulnerability of cultivars and hybrids to a multitude of insect pests and diseases. The presence of the toxic proteins ricin and hyperallergenic *Ricinus communis* agglutinin (RCA) in the endosperm restricts the use of deoiled seed cake as cattle feed. Due to this crop's low genetic diversity, genetic engineering can be an efficient approach to introduce resistance to biotic and abiotic stresses as well as seed quality traits. Recently, castor oil gained attention as a sustainable second-generation feedstock for biojet fuel that would reduce carbon dioxide emissions. Because of a growing interest in castor oil as a biofuel and the presence of the powerful toxin ricin in its seed, metabolic pathways and regulatory genes involved in both oil and ricin production have been analyzed and characterized. Genetic engineering of castor bean offers new possibilities to increase oil yield and oxidative stability, confers stress tolerance, and improves other agronomics traits, such as reduced plant height to facilitate mechanical harvesting. However, difficulties in tissue culture-based regeneration and poor reproducibility of results are major bottlenecks for genetic transformation of castor bean. Despite advances in tissue culture research over the past four decades, direct or callus-mediated adventitious shoot regeneration systems that are genotype-independent remain a much sought-after goal in castor bean. Genetic transformation attempts to develop insect-resistant and ricin-free transgenic castor bean lines have been based on shoot proliferation from meristematic tissues. This chapter describes new transformation methods under development and the progress achieved so far in genetic engineering of castor bean for agronomically desirable attributes.

14.1 Introduction

Castor bean (*Ricinus communis*) is a hardy crop easily adaptable for tropical, subtropical, and temperate regions. However, it is reported to be vulnerable to a wide array of pests and diseases that affect vegetative and reproductive tissues during the entire crop cycle. In addition, it is susceptible to drought, requiring at least 600–700 mm of rainfall per year, and it is sensitive to salinity, poorly drained soils, and frost, particularly during early growth stages. Genetic

M. Sujatha (✉) · M. Tarakeswari
ICAR-Indian Institute of Oilseeds Research, Rajendranagar, Hyderabad 500030, Telangana, India
e-mail: mulpuri.sujata@icar.gov.in

C. Kole and P. Rabinowicz (eds.), *The Castor Bean Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97280-0_14

improvement of castor bean has been mostly limited to the exploitation of natural genetic variability and focused on the selection for high yield, desirable branching types, sex morphotypes, nonshattering capsules, disease resistance (e.g., Fusarial wilt), and high seed oil content.

Extensive cultivation under high inputs and without proper scientific management and crop rotation has made castor bean vulnerable to a number of biotic and abiotic stresses. About 100 species of castor bean insect pests have been recorded at different phenological stages of the crop. Among those pests, castor semilooper (*Achaea janata*), capsule borer (*Conogethes punctiferalis*), tobacco caterpillar (*Spodoptera litura*), red hairy caterpillar (*Amsacta albistriga*), jassids (*Empoasca flavescens*), and white fly (*Trialeurodes ricini*) cause considerable damage (Lakshminarayana and Raoof 2005). Seed yield losses are estimated to be in the range of 35–50%, depending on the crop growth stage and the pest type. Severe outbreaks of castor semilooper and red hairy caterpillar during early crop growth that resulted in complete crop loss have been reported. Castor bean can tolerate up to 25% leaf damage without affecting seed yield, while damage caused to spikes and capsules often leads to significant yield losses. Diseases, such as wilt (*Fusarium oxysporum* spp. *ricini*), root rot (*Macrophomina phaseolina*), bacterial blight (*Xanthomonas ricini*), botrytis gray rot (*Botrytis ricini*), and seedling blight (*Phytophthora colocasiae*), are considered important (Kolte 1995; Lakshminarayana and Raoof 2005).

New sources of abiotic and biotic stress tolerance are in constant demand by breeders. *Ricinus* is considered to be a monotypic genus, and *R. communis* is the lone species encompassing the many polymorphic types known around the world (Weiss 1983). Several of these types were designated as species (*R. communis, R. macrocarpus, R. microcarpus*), but they are intercrossable and fertile and therefore not true species. All castor varieties that have been investigated cytologically are diploids with a $2n$ number of 20 chromosomes (Richharia 1937), and the species is a secondary balanced polyploid with a basic number of $x = 5$ (Singh 1976).

Successful breeding for yield stability is thus limited by a lack of exploitable genetic variability of productivity and disease and pest resistance traits among castor bean germplasm. Breeders have to resort to alternative approaches like mutations, wide (intergeneric) hybridization, and biotechnology for the creation of genetic variability and incorporation of desired traits into castor. The effect of irradiation on castor bean seeds and seedlings has been the subject of a number of studies that aimed at identifying mutants with specifically required characters. The importance of induced mutations in castor has been demonstrated in the development of productive semidwarf lines with high yield potential, early maturing mutants, and sex expression variants (Kulkarni and Ankineedu 1966; Ankineedu et al. 1968). However, mutants that are resistant or tolerant to biotic stresses or mutants that have reduced seed toxicity have not been obtained by radiation mutagenesis so far.

In several crops, wild relatives have often been used in breeding programs as they constitute important reservoirs of genetic variability (Harlan 1976). Morphologically, the genus *Jatropha* ($2n = 22$) resembles *Ricinus* and has attracted interest because it possesses useful traits not found in castor. Although both genera, as well as *Manihot* (cassava), belong to the Euphorbiaceae family, successful introgression of some of the desirable genes of *Jatropha* and *Manihot* into castor bean remain virtually untapped, because of the incompatibility of intergeneric crosses between these genera (Reddy et al. 1987a; Sujatha 1996; Gedil et al. 2009).

In addition to castor bean's susceptibility to insect pests and the presence of toxic proteins in the endosperm, its perennial and indeterminate growth habit are further challenges for breeding early maturing varieties that are amenable for mechanical harvesting.

Castor bean produces various industrially useful fatty acids, and renewed interest in this crop is due to its potential to produce biofuels and epoxy oil through genetic engineering. Because epoxy oil and castor oil have very similar chemical structures, only minor genetic

modifications would be necessary to synthesize epoxy oil in castor bean. Genes involved in triacyl glycerol assembly and fatty acid metabolism, such as stearoyl-acyl carrier protein desaturases (*SAD2, SAD3, SAD4*), diacylglycerol acyl transferases (*DGAT, DGAT2*), enolases, lipid transfer proteins, fatty acid desaturases, hexose transporters, 6-acyl-Sn-glycerol-3-phos phate acyltransferases (*LPAAT2, LPAAT4, LPAAT5*), oleate desaturases (*FAD2*), 3-keto-acyl-ACP synthases (*KASI, KASII, KAS III*), oleosins (*Ole1* and *Ole2*), and lipid transfer proteins (Weig and Komor 1992), can increase oil content and modify the plant's fatty acid profile. Genes coding for the highly toxic protein ricin is also important for the development of cultivars with reduced toxicity. Such genes could be used to generate improved cultivars by genetic engineering. Furthermore, castor bean is an excellent model system for seed development due to its large-sized seed with high lipid and protein content (Kermode et al. 1989), as well as to study sucrose and amino acid uptake mechanisms (Komor 1977; Robinson and Beevers 1981) and nutrient transport through phloem (Orlich and Komor 1992) and xylem (Schobert and Komor 1992).

With the availability of a draft sequence of the castor bean genome, most of the genes underlying oil metabolism, ricin, agglutinin, and disease resistance have been identified (Chan et al. 2010). Malik (2013) suggested the possibility of combining genes for heat tolerance, drought tolerance, and resistance to insect pests and diseases to improve tolerance and high oil yield in castor. Also, new genome editing tools could be applied to eliminate ricin from seed meal and make epoxy oil in castor bean.

One popular strategy to mitigate insect pest susceptibility is the introduction of suitable resistance genes through transformation (Sharma et al. 2000). The 110-fold hectare increase of cultivation of transgenic crops in two decades between 1996 and 2016 makes genetic engineering the fastest adopted crop technology in agriculture (James 2016). Of the 185 million ha of cultivated biotech crops during 2016 globally, 12.5% were occupied by insect-resistant crops, and 41% were crops with stacked double and triple traits.

During the past three decades, intensive studies on plant regeneration and transformation have led to the production of transgenic plants in many crop plants. The introduction of foreign genes by genetic engineering techniques often requires an efficient in vitro regeneration system for the desired plant species. Such a system must be rapid, reliable, and applicable to a broad range of genotypes. However, facile techniques for tissue culture, regeneration, and gene transfer have not been established for castor bean, despite decades of research. The protocols for stable transformation that are in place for castor bean utilize meristematic tissues, which could lead to the recovery of chimaeras, but there is potential for genetic improvement of this species through in vitro techniques and biotechnological tools.

14.2 Castor Bean Tissue Culture

The major hindrance for genetic transformation of several crop plants is the availability of a reproducible system for plant regeneration. In many cases, transformation methods target young apical meristems or other tissues that ultimately give rise to gametes (Birch 1997). In castor bean, the development of an efficient and highly reproducible system for tissue culture and regeneration remains a major challenge for genetic transformation experiments. Castor bean proved to be highly recalcitrant to in vitro manipulations, which is a major impediment for the development of transgenic lines (Reviewed by Sujatha et al. 2008). Early attempts at establishing castor bean tissue culture were carried out during 1960s using endosperm tissue mostly because its large endospermic seeds facilitated in vitro culture. However, these experiments resulted in continuously growing tissue that did not undergo organogenic differentiation. Tissue culture studies of castor bean undertaken during 1980s aimed at obtaining whole plantlet regeneration from seedling tissue. These experiments were conducted on young seedlings, but regeneration of complete plants was rather limited.

Plant regeneration was mainly derived from pre-existing meristematic centers (Athma and Reddy 1983; Sangduen et al. 1987; Reddy and Bahadur 1989; Molina and Schobert 1995; Alam et al. 2010; Li et al. 2015a), and a maximum of 40 and 47 shoots from embryo axes and shoot tip explants, respectively, was reported (Sujatha and Reddy 1998). Callus-mediated shoot regeneration from hypocotyl explants, young stem segments, leaves, and cotyledonary leaves have been reported, but morphogenic differentiation was sporadic, unreproducible, and with very low frequency of shoot regeneration, showing only 1–5 shoots per responding explant (Reddy et al. 1987b; Genyu 1988; Bahadur et al. 1992; Sarvesh et al. 1992). Subsequently, Ahn et al. (2007), Sujatha and Reddy (2007), and Ganesh Kumari et al. (2008) reported relatively higher shoot induction frequencies from seedling explants, obtaining 22–24 shoots per explant. They used growth adjuvants and amino acids to improve caulogenic ability. Studies by Zhang et al. (2016) suggested that a 6-day dark treatment significantly increased the average number of shoots (37) per explant when cultured on medium supplemented with 0.3 mg/L thidiazuron (TDZ). Li et al. (2015b) compared WUSCHEL (*WUS*) gene expression in castor bean tissues under different inducing conditions and optimized a regeneration system based on embryogenic cell induction. These studies showed a positive correlation between *WUS* gene expression and embryogenic cell induction, which had direct influence on the rate of shoot budding. Nevertheless, the reproducibility of these methods across laboratories and different genotypes has to be ascertained as very few genotypes were used in most of these studies.

Meristematic explants were tested in castor bean, and embryo axes showed high proliferative ability as compared to shoot apices and nodal explants (Sujatha and Reddy 1998). Shoot proliferation rates reported in this investigation were very high and were 40 and 81.7 from embryo axes and 46.7 and 22.0 shoots per explant from shoot tips on induction and subculture media, respectively. The study demonstrated a carryover effect of TDZ in embryo axes from mature seeds for 2–3 subculture cycles. TDZ-habituated embryo axes tended to produce shoots continuously, reaching up to 81.7 shoots per explant when cultures were transferred from medium with 5.0 mg/L TDZ to medium supplemented with 0.5 mg/L benzyladenine (BA). The elevated proliferation rate of zygotic embryos on medium supplemented with TDZ paved the way for genetic transformation studies in castor bean (Sujatha and Sailaja 2005; Malathi et al. 2006).

Genotypic differences were not significant in shoot proliferation experiments on medium supplemented with TDZ. Studies of Ahn et al. (2007) were in agreement with those of Sujatha and Reddy (1998), which showed a threefold increase in cytokinin activity of TDZ relative to BA for promoting shoot regeneration from mature zygotic embryos. Pretreatment of embryo axes in the dark for 7 days increased the number of shoots regenerated per explant by 82 and 36% with TDZ and BA, respectively (Ahn et al. 2007). Likewise, preincubation of cotyledon explants from mature seeds cultured on medium with 5 μM TDZ in the dark for 7 days resulted in a maximum of 25 shoots per explant (Ahn and Chen 2008). In both studies, histological analysis indicated adventitious origin of the shoots from the meristematic region, including the shoot apex and the cotyledonary leaf axil. TDZ showed a beneficial effect on adventitious shoot regeneration as well. Ahn et al. (2007) reported a high frequency of shoot regeneration from hypocotyls on medium supplemented with 0.25 mg/L TDZ and 4.5 mg/L BA. Zalavadiya et al. (2014) reported shoot regeneration from hypocotyl-derived callus on medium supplemented with 0.5 mg/L kinetin and 0.25 mg/L BA, which was reproducible across genotypes and in different laboratories. However, this protocol suffers from lack of rooting of the regenerated shoots even after 30 days in culture.

Tissue culture research conducted so far in castor bean has revealed a strong recalcitrance to in vitro culture and identified the development of an adventitious shoot regeneration system as a major challenge. The observed sporadic shoot regeneration indicates the presence of a few morphogenic cells interspersed within

nonmorphogenic tissues. Although morphogenically competent cells are more likely to be found in juvenile than in mature tissues of flowering plants, suppression of competent cells by overgrowth of noncompetent cells may have been the reason for the failure to isolate a competent line in castor bean. The occasional appearance of shoots could also be due to the activation of recalcitrant calli to undergo caulogenesis caused by rare exogenous and endogenous inductive stimuli. Hence, a large number of genotypes and growth regulator combinations need to be assessed for determining caulogenic ability and to understand recalcitrance in castor bean tissues in vitro. Until an efficient and reproducible system for plant regeneration is developed, genetic transformation of castor will be a meristem-based shoot proliferation system.

14.3 Genetic Transformation Methods for Castor Bean

Castor bean has been reported to be susceptible to crown gall disease caused by *Agrobacterium tumefaciens* (Lippincott and Haberlein 1965), which is broadly used for plant transformation due to its capacity to introduce DNA into plant cells. Genetic transformation methods tried for castor include *Agrobacterium*-mediated transformation and particle gun bombardment methods with each of these methods having its own advantages and disadvantages.

14.3.1 *Agrobacterium*-Mediated Method

Owing to the poor caulogenic ability of in vitro cultured tissues in castor, transformation was carried out with tissues with pre-existing meristems like decotyledonated embryo axes or using alternate methods bypassing tissue culture-based regeneration such as floral bud transformation or *in planta* transformation.

Agrobacterium-mediated transformation is one of the most efficient methods for gene transfer and takes advantage of the naturally evolved crown gall-inducing mechanisms of DNA present in *A. tumefaciens*—a ubiquitous gram-negative soil pathogen. The uniqueness of this soil microorganism lies in its capability of transfer of DNA to eukaryotic cells. Although it is genotype-dependent, it is the most preferred method of gene transfer owing to its simplicity, cost-effectiveness, and generation of stable transformants with a single or few copies of the inserted DNA fragment, which can be relatively large and with defined ends (Gelvin 2003).

The *Agrobacterium*-mediated transformation process involves a number of steps which include: (a) isolation of the genes of interest from the source organism; (b) development of a functional transgenic construct harboring the gene of interest along with promoters to drive expression and marker genes (selectable such as antibiotic or herbicide markers or reporter genes like *GUS* or *GFP*) to facilitate tracking of the introduced genes in the host plant; (c) insertion of the transgene into the Ti plasmid; (d) introduction of the T-DNA-containing plasmid into *Agrobacterium*; (e) cocultivation of plant cells with transformed *Agrobacterium* to allow transfer of T-DNA into plant chromosome; (f) regeneration of the transformed cells into putative transformants; (g) confirmation of integration and expression of the introduced transgene through molecular analysis; and (h) assessing the trait performance through appropriate assays at lab, greenhouse, and field level.

In planta transformation methods like floral dip method that have been developed for *Arabidopsis* are incredibly simple and extended to transformation of several other crops. In floral bud transformation method of castor, the flower buds of an intact plant are wounded and an *Agrobacterium* strain carrying the transgene is infiltrated into the wound (US Patent No 6.620.986). The flower buds are allowed to set seed, and those seeds that harbor the transgene (which typically includes a selectable marker) are grown under the selective medium. However, in this method, transformation of the germline (so as to enable the progeny carry the transgene) is extremely difficult because the floral initials are

usually at different stages of differentiation along the racemes. Another method is to prick the shoot apical meristems of 10–15-day-old seedlings with fine needles followed by immersion in *Agrobacterium* suspension and transfer to soilrite in bottles for a week (Kumar et al. 2011). Subsequently, the infected seedlings were transferred to pots and grown to maturity in the greenhouse. Transformed lines will be selected in the next generation through antibiotic/herbicide screening. The advantage of *in planta* method lies not only in the simplicity and reliability but also in obtaining genetically uniform (nonchimeric) transformed progeny as somaclonal variation associated with tissue culture and regeneration is minimized. Although the method is simple, it requires extensive analysis of the T_1 plants for identification of plants carrying the transgene.

14.3.2 Particle Gun Bombardment

The development of particle gene gun has revolutionized DNA transfer technology, bypassing limitations imposed by *Agrobacterium* host specificity and cell culture constraints, and has allowed the engineering of almost all the major crops including formerly recalcitrant cereals, legumes, woody species, and other organisms. It is a physical method in which rapidly propelled tungsten or gold particles coated with DNA are blasted to deliver biologically active DNA into plant cells. Direct gene transfer through particle gun bombardment is generally genotype-independent, facilitating multiple gene cotransformation, but the success depends on several physical parameters and is often reported to result in transgenics with multiple copies of the transgene.

Regardless of the method employed, most transformation studies in castor bean have targeted meristematic tissues, which are prone to result in chimeras. Because of its highly proliferative ability on medium supplemented with a strong cytokinin like TDZ or a combination of TDZ and BA, castor bean embryo axes have been the most used tissue for transformation, regardless of the method. Success depends on the precise targeting of the introduced DNA to the wounded meristematic tissues and the conditions (*Agrobacterium* culture density, period of cocultivation, use of *vir* gene inducers like acetosyringone and other phenolic compounds) for enhancing the virulence (the transfer of T-DNA from *Agrobacterium* to its host).

McKeon and Chen (2003) obtained 12 genetically engineered castor bean plants through vacuum infiltration of wounded flower buds in *Agrobacterium* suspension (US Patent No 6.620.986). Sujatha and Sailaja (2005) described the first successful attempt at development of a stable transformation system for castor bean using decotyledonated embryo axes. In this method, embryo axes following cocultivation with *Agrobacterium* were subjected to expansion and proliferation on Murashige and Skoog (1962) (MS) medium supplemented with 0.5 mg/L TDZ followed by three cycles of selection on medium with 0.5 mg/L BA and increasing concentrations of hygromycin (20–40–60 mg/L) or kanamycin (50–100–200 mg/L) depending on the selectable marker (*hpt, npt* II). Selected shoot clusters were transferred to medium with 0.5 mg/L BA for proliferation and 0.2 mg/L BA for shoot elongation. Elongated shoots were rooted on half-strength MS medium supplemented with 2.0 mg/L NAA. By employing this protocol, a primary transformant was obtained within 5 months from cultured embryo axes with an overall transformation efficiency of 0.08%. As the protocol does not involve an intervening callus phase, no abnormal phenotypes were reported through this procedure. A similar shoot proliferation method with minor modifications was followed for direct gene transfer using the particle gun bombardment method achieving a transformation efficiency of 1.4% (Sailaja et al. 2008). Transformation frequencies of castor bean reported to date are generally low (Table 14.1).

Table 14.1 Genetic transformation studies in castor bean

Transformation method	Gene inserted	Explant	Trait	Transformation frequency (%)	References
Agrobacterium-mediated	*Gus*	Wounded flower buds	Transformation protocol	–	McKeon and Chen (2003)
Agrobacterium-mediated	*Gus, hpt, npt*	Embryo axis	Transformation protocol	0.08	Sujatha and Sailaja (2005)
Agrobacterium-mediated	*cry1Ab*	Embryo axis	Insect resistance (semilooper)	0.42	Malathi et al. (2006)
Particle gun method	*Gus, hpt, npt*	Embryo axis	Transformation protocol	1.4	Sailaja et al. (2008)
Agrobacterium-mediated and particle gun methods	*cry1EC*	Embryo axis	Insect resistance (semilooper and Spodoptera)	0.82 and 0.69, respectively	Sujatha et al. (2009)
Agrobacterium-mediated	*Chitinase – Chi 1*	Cotyledonary node	Fusarium wilt	–	Ganesh Kumari (2010)
Agrobacterium-mediated *in planta* methods	*cry1ACF*	Two-days old seedlings	Insect resistance (Spodoptera)	1.4	Kumar et al. (2011)
Agrobacterium-mediated	*Ricin toxin A chain*	Cotyledonary node	Reduced seed toxicity	–	Chen et al. (2013)
Agrobacterium-mediated	*Ricin toxin A chain*	Epicotyl	Reduced seed toxicity	–	Li et al. (2014)
Agrobacterium-mediated	*SbNHX I*	Embryo axis	Salt stress	2.8–5.9	Patel et al. (2015)
Agrobacterium-mediated	*TFL1*	–	Determinate and early maturity	–	Peles et al. (2017)
Agrobacterium-mediated	*Ricin toxin A chain*	Embryonic axis	Reduced seed toxicity	0.85	Sousa et al. (2017)

14.4 Genetic Transformation of Castor Bean with Insect Resistance Genes

Before embarking on a genetic engineering program to address a pest problem, it is imperative to identify suitable insect resistance genes, which could be introduced into castor bean against major pests. Several candidate genes, such as crystal protein (*Cry*) genes of *Bacillus thuringiensis* (*Bt*) produced during the sporulation stage, vegetative insecticidal *Bt* proteins (VIPs) induced during the vegetative stage, proteinase inhibitors, lectins, α-amylase inhibitors, insect chitinases, and novel genes of plant origin, can be deployed into crop plants for imparting protection against insect pests. However, the most commonly used and commercially exploited insect resistance genes are the *Bt Cry* genes. Insecticidal δ-endotoxins of *B. thuringiensis* have acquired great significance because of their specificity to target pests, nontoxicity to humans and beneficial insects, toxicity at low concentration, and environment-friendly nature. *B. thuringiensis* var. *kurstaki* strains produce several lepidopteran toxic proteins such as Cry1Aa, Cry1Ab, Cry1Ac, Cry1IA, and Cry1B. Information on the reaction of the major lepidopteran pests attacking castor bean to Cry

proteins in the toxin specificity database (http://www.glfc.forestry.ca/bacillus/web98.adb) is limited. Experiments were undertaken at the Indian Institute of Oilseeds Research to assess the efficacy of various purified crystal *Bt* proteins which are lepidopteran-specific against major defoliators of castor bean (Lakshminarayana and Sujatha 2005; Sujatha and Lakshminarayana 2005).

Bioassays against neonate larvae of *Achoea janata* (castor semilooper), *S. litura*, *Spilosoma obliqua*, and *Euproctis fraterna* using Cry toxins (Cry1Aa, 3A, 2B, 1C, 2A, 1E, 1Ac, 1F, 9A, 1Ab) at concentrations ranging from 4 to 1500 ng/cm^2 were done using leaf paint assays. With regard to semilooper, the Cry proteins 1Aa, 1Ab, 1E, and 2A were found to be the most effective, resulting in 100% mortality within 48 h, while the other proteins gave nil or delayed mortality at the highest concentrations tested. Among the effective proteins, Cry1Aa was found to be superior to other proteins in giving early mortality even at lower concentrations (125 ng/cm^2) (Sujatha and Lakshminarayana 2005). In the case of Spodoptera, none of the proteins gave 100% mortality even after 96 h of treatment at the highest concentration tested (1500 ng/cm^2), except for Cry1Aa, which gave 50% mortality at 1500 ng/cm^2. Increasing the concentration of the proteins up to 3000 ng/cm^2 also failed to cause larval mortality. However, feeding cessation in terms of low larval weight was recorded in treatments with Cry1Aa and Cry 1Ab (Lakshminarayana and Sujatha 2005). The Cry proteins 1Aa, 1E, 1Ab were found to be effective against *S. obliqua*, while 1Ac, 1Aa were effective against *E. fraterna.*

Due to Cry1Aa protein's efficacy against all tested pests, genetic transformation of castor bean has been initiated using the *cry1Aa* gene modified for plant codon usage. *S. litura*, which is a polyphagous lepidopteran insect that affects castor bean, is tolerant to most of the known δ-endotoxin proteins. Hence, a hybrid Cry1Ea/Cry1Ca δ-endotoxin protein called Cry1EC was developed by replacing amino acid residues 530–587 in the low-activity Cry1Ea protein with a highly homologous 70-amino acid region of Cry1Ca in domain III (Singh et al. 2004). The soluble Cry1EC protein produced with an expression vector was fourfold more toxic to the larvae than Cry1Ca, the most effective δ-endotoxin against *Spodoptera* sp. This Cry1EC hybrid endotoxin conferred complete protection against *S. litura* when deployed in tobacco and cotton. Based on these studies, the genes *cry1Aa* and *cry1Ab* have been selected to confer protection against major foliage feeders of castor bean, and transgenic castor bean lines harboring the chimeric genes *cry1EC* (Sujatha et al. 2009) and *cry1AbcF* (Kumar et al. 2011), as well as the synthetic δ-endotoxin gene *cry1Ab* (Malathi et al. 2006), have already been generated.

The vector-mediated and direct gene transfer methods (Sujatha and Sailaja 2005; Sailaja et al. 2008) were employed for transformation of castor bean cv. DCS-9 using appropriate vectors containing the *cry1Aa* (Sujatha M, unpublished) and the *Bt* fusion gene *cry1EC*, driven by an enhanced 35S promoter (Sujatha et al. 2009). With the *cry1EC* gene, 81 and 12 putative transformants were obtained following selection on hygromycin and kanamycin, respectively. The integration and inheritance of the introduced genes were demonstrated up to T_4 generation by PCR and Southern blot analysis. Field bioassays against *S. litura* and castor semilooper, conducted for eight transformation events in T_1–T_4 generations under net confinement, lead to identification of promising events conferring resistance to the two major defoliators (Fig. 14.1).

A similar procedure was used for the production of semilooper-resistant transgenic castor bean that expresses a synthetic δ-endotoxin *cry1Ab* gene driven by CaMV (cauliflower mosaic virus) 35S promoter (Malathi et al. 2006). The construct harboring the insect resistance gene carried the herbicide resistance gene (*bar*) for selection of putative transformants. The presence of the introduced gene, its stable integration, expression, and inheritance was confirmed through PCR, Southern blot analysis, ELISA, and progeny tests. The transformed plants showed a Cry1Ab protein concentration between 0.23 and

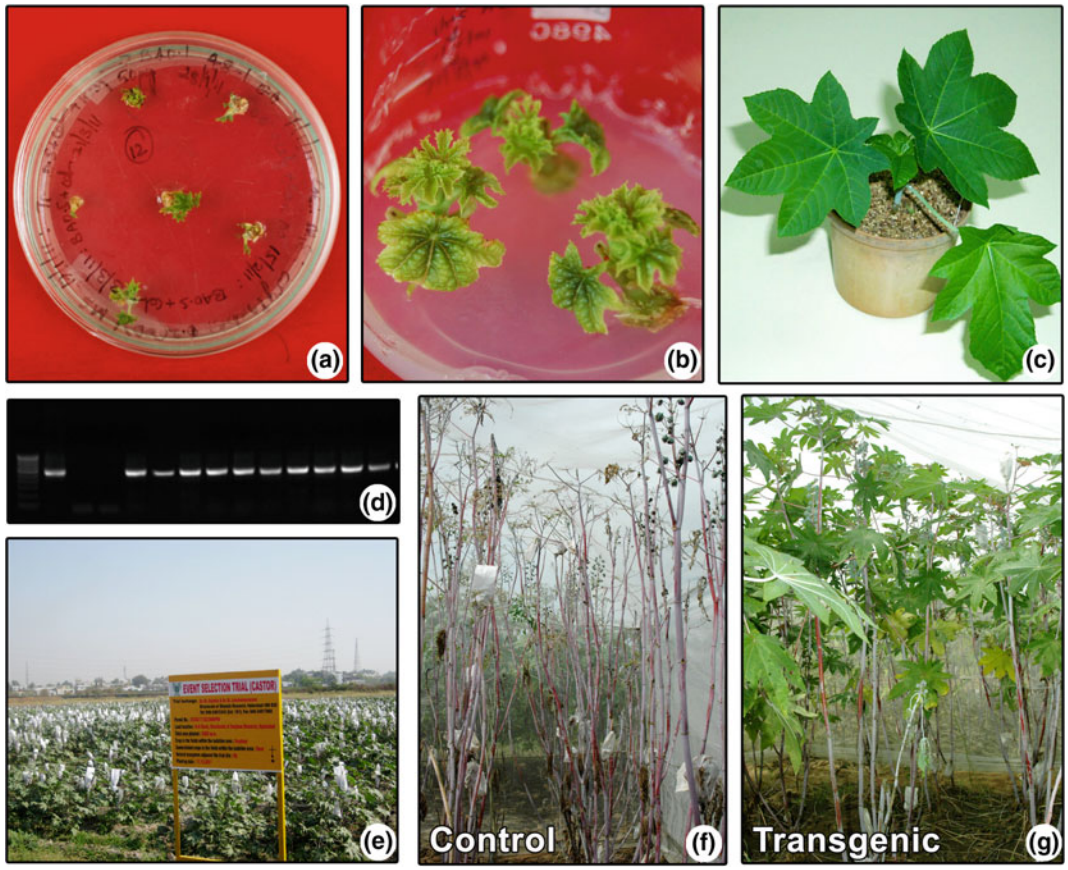

Fig. 14.1 Genetic transformation in castor for insect resistance. **a** Selection of putative transformants on kanamycin; **b** elongation of putative transformed shoots; **c** acclimatization of transformed shoots; **d** identification of homozygous line through PCR analysis; **e** event selection trial for generation advancement and insect bioassays, **f** reaction of control (untransformed) plants showing complete defoliation; **g** reaction of the transgenic lines showing less foliar damage compared to control plants

0.47 ng/mg in the leaf tissue. Biological assays in the laboratory with neonate larvae of semilooper showed a marked feeding inhibition associated with reduced larval growth and substantial mortality (88.9–97.3%) on different primary transformants as compared to the untransformed control plants (13.9%). This and the above study (Sujatha et al. 2009) used the castor bean genotype DCS-9 (Jyoti), which is a cultivated commercial variety and a parental line for the hybrid DCH-177, and therefore, it will not be necessary to transfer the transgene between experimental and commercial varieties.

Kumar et al. (2011) used the *in planta* transformation procedure through *Agrobacterium*-mediated transfer for introduction of the *cry1AbcF* gene against Spodoptera. In this method, two-day-old seedlings were infected with *Agrobacterium* strain EHA105/pBin*Bt*8 harboring the *cry1AcF* gene and transformed rooted shoots were established in the greenhouse. Subsequently, T_1 generation seedlings were screened on 300 mg/L kanamycin to select putative transformants and five highly expressing transgenic lines were identified through molecular and expression analyses. These transgenic lines showed high mortality of semilooper and the stability of the transgene up to T_2 generation was successfully demonstrated.

The studies dealing with evaluation of transgenics for conferring protection against the two major castor bean foliage feeders were mostly confined to laboratory bioassays except for the study dealing with *cry1EC* transgenics (Sujatha et al. 2009) where whole plant assays were conducted. In the case of *cry1Ab* transformants, T_0 plants were bioassayed against semilooper (Malathi et al. 2006), while T_1 and T_2 transformants harboring the *cry1AbcF* gene were assessed for their reaction to Spodoptera in detached leaf assays in the laboratory (Kumar et al. 2011). The transformants harboring the *cry1EC* gene were tested for their reaction to castor bean semilooper and Spodoptera under laboratory conditions using detached leaves and at whole plant level under net-contained field conditions (Sujatha et al. 2009). Insect bioassays coupled with ELISA studies revealed variation in the plants' pest response and in protein expression levels, depending on transformant generation, crop stage, and growing conditions.

The identification of transformation events that result in plants bestowed with resistance to foliage feeders is just a beginning of generating pest-resistant castor bean lines. It is then necessary to determine toxin expression levels in different transformants and developmental stages. It is also essential to assess the toxicity of the candidate genes to beneficial insects. Eri silkworm (*Samia cynthia ricini*) larvae are reared on castor leaves for silk production. Kumar et al. (2016) observed high toxicity of purified Cry1Aa and Cry1Ab crystal proteins of *B. thuringiensis* against *eri* silkworm.

Traditional capsule borer, *Conogethes (Dichocrosis) punctiferalis* which used to be at low key, has increased in India causing up to 50% capsule damage in castor (Duraimurugan et al. 2015). The management of defoliators like semilooper and *Spodoptera* is relatively easy as the plant has the ability to tolerate a certain degree of defoliation, and these pests have potential natural enemies and are susceptible to a wide range of insecticides. Contrary to this, the management of capsule borer is rather difficult as the pest attacks the inflorescence and growing capsules, which directly translates into yield reduction. Capsule borer is less susceptible to chemical insecticides and hence is difficult to manage. Suitable management practices other than chemical control are not available. For capsule borer, the major challenge lies in optimizing a rearing technique prior to testing the effective agents and identification of candidate gene(s) for introduction through genetic engineering tools.

14.5 Transgenic Castor Bean for Abiotic Stress Tolerance

The first generation of genetically modified crops that have been cultivated since the 1990s in a number of countries harbor single genes that confer tolerance to herbicides or insect attacks (www.isaaa.org). Second-generation transgenic crops with stacked genes or traits that allow crops to tolerate environmental stresses such as drought, cold, salt, heat, or flood are now being tested in contained field trials around the world. In the case of castor bean, transgenic development is still in its infancy and there are no commercially released lines as of today. With regard to abiotic stresses, an Israeli biotechnology company—Rahan Meristem (http://www.rahan.co.il)—is developing castor bean transformation protocols to generate salinity and drought resistance using genes from unicellular algae, which are expected to improve the performance of castor bean in low rainfall climates.

In order to enhance salt stress in castor, Patel et al. (2015) introduced the vacuolar Na^+/H^+ antiporter gene *SbNHX1* into castor bean with an improved *Agrobacterium*-mediated transformation method that uses spermidine (1 mM) and acetosyringone (200 μM). Stable integration and expression of the transgenes were confirmed by PCR and Southern blot hybridization. Physio-biochemical analyses such as quantification of chlorophyll, water, proline, malondialdehyde (MDA), Na^+ and K^+, as well as

measurements of K^+/Na^+ ratios, and electrolytic leakage under varying NaCl concentrations showed that ectopic expression of *SbNHX1* enhances salt tolerance in transgenic plants by modulating physiological process.

14.6 Introduction of Other Agronomically Desirable Traits into Castor Bean

14.6.1 Altering Plant Architecture

Flowering plants exhibit either determinate or indeterminate type of inflorescence architecture. Determinate inflorescences are ideal for uniform and high-density planting, and it is amenable for mechanization and combined harvesting. The inflorescence of cultivated castor bean, including ruling hybrids and varieties, is predominantly indeterminate making the crop less amenable for mechanical harvesting.

Castor bean genotypes with low tendencies to branch have been selected using the pedigree method, but successive rounds of self-pollination of these cultivars reduced plant vigor (Baldanzi and Pugliesi 1998). According to Severino et al. (2012), the selection for short and nonbranching castor bean plants is usually difficult due to their high genotype versus environment interaction. According to Brigham (1980), the transition from the vegetative to the reproductive stage of the main stem corresponding to the formation of the main raceme leads to production of axillary shoots. From these shoots, the second and subsequent order racemes originate and the sequence of development continues as long as the plants remain alive and healthy, thus producing an indeterminate growth habit.

Because terminal flowers are produced in determinate types, two important regulators of the flowering pathways, viz. FLOWERING LOCUS T (FT) and TERMINAL FLOWER 1 (TFL1) which function in diverse signaling pathways have been studied. FT promotes the transition to reproductive development and flowering, while TFL1 represses it. A loss of function of the TFL-1 gene has been associated with a determinate phenotype and early flowering. An early flowering and determinate castor bean inbred line have been developed through transgenic downregulation of TFL1 activity. The resulting plants exhibited stable vigor for at least five generations (Peles et al. 2017).

14.6.2 Seed Traits

The demand for castor oil for expanded industrial applications in manufacturing of biodegradable products and as a source of biofuel is increasing. Consequently, there is a need for developing nontoxic, low-ricinine, and nonallergenic castor bean for safe utilization of the seed meal for animal feeding besides increasing the oil content and modification of oil quality for varied purposes.

14.6.3 Toxin-Free Castor

The major toxic constituents of castor bean seed meal are the two-chained polypeptide ricin, the alkaloid ricinine, the four-chained polypeptide *R. communis* agglutinin (RCA), and the allergenic protein polysaccharide CB-1A (castor bean allergen). Due to the presence of these deadly natural poisons and allergens, castor bean's high-protein meal is only used as fertilizer or it is incinerated. Enhanced expression of ricin, 2S albumin, and ricinoleate production start at 26 days after pollination (Chen et al. 2007). Ricin is a type II ribosome inactivating protein (RIP) and is deadly when inhaled, ingested, or injected (Lord et al. 1984; Hartley and Lord 1993). Chemical detoxification procedures are available, but the energy input needed to treat the deoiled seed meal to destroy the toxic proteins and allergens limits the economic competitiveness of such procedures. A major objective of castor bean genetic transformation is to develop ricin-free varieties that are safe for growers and seed processors while increasing the value of the high-protein meal remaining after castor oil extraction. This problem is being tackled through a multipronged approach. Conventional and

biotechnological approaches are being attempted to develop cultivars with reduced levels of toxin. Traditional breeding methods involving crosses of the high-yielding dwarf line 'hale' with two lines that show low levels of seed toxins resulted in the development of a dwarf plant line with a 70–75% reduction in ricin and RCA content (Auld et al. 2003). Gene silencing approaches to reduce the levels of toxins/allergens through crop genetic engineering have been successful. The ricin gene has been isolated and sequenced, and analyses of the transcription of ricin, allergen, and numerous lipid biosynthetic enzymes during seed development led to the identification of promoters that can be used to express genes that suppress toxin and allergen production. A combination of breeding and transgenic approaches is expected to produce castor plants that have potential for 99.9% reduction in ricin content in segregating populations (Auld et al. 2001). However, transgenic strategies through antisense silencing resulted in limited toxin reduction as ricin production is controlled by multiple genes. Studies by Pinkerton et al. (1999) indicated that ricin concentration is controlled by multiple genes and is also influenced by the environment.

The biotech company Arcadia Biosciences has focused on 'knocking out' toxic genes using classical mutagenesis and tilling (targeting induced local lesions in genomes), which identifies single nucleotide polymorphisms (SNPs) in known genes responsible for ricin production (www.arcadiabio.com). Ricin was initially considered to belong to a lectin multigene family composed of eight members (four A and four B chains). Based on this information, research was undertaken at Mississippi State University, USA, USDA-ARS, Albany, USA, and the Indian Institute of Oilseeds Research, Hyderabad, India, to genetically engineer castor for inhibiting ricin expression and/or production. Later on, the annotation of the draft castor bean genome sequence identified 28 putative genes in the ricin family, including potential pseudogenes or gene fragments (Chan et al. 2010).

RNA interference (RNAi) technology has been exploited to silence the ricin genes in castor bean seed endosperm to develop cultivars with reduced toxicity (Chen et al. 2013; Li et al. 2014; Sousa et al. 2017). Sousa et al. (2017) cloned two copies of a 460 bp fragment from the ricin toxin A-chain gene (RTA) in concatenated sense and antisense orientations under the control of the 35S CaMV promoter in order to generate an intron-hairpin RNAi cassette (Δ*ricin*). The cloning vector also contained the reporter *gus* gene and a mutated *ahas* gene from Arabidopsis, which confers tolerance to the imidazolinone class of herbicides. Results indicated effective silencing of the ricin genes in the transgenic plants and ricin proteins could be not detected by ELISA. Protein extracts from transgenic seeds lacked hemagglutination activity and were not toxic to rat intestine epithelial cells or Swiss Webster mice. Chen et al. (2013) also succeeded in obtaining transgenic castor bean plants containing a 762 bp fragment of the RTA by *A. tumefaciens*-mediated transformation of cotyledonary nodes. Li et al. (2014) selected a 351-bp fragment of the RTA to produce an RNAi construct that was introduced into castor bean through *Agrobacterium*-mediated transformation of epicotyls. Semiquantitative analysis showed an apparent decrease in the RTA expression level, further suggesting that RNAi is an effective approach for reducing ricin mRNA expression. The biodetoxified castor bean cake which is very rich in valuable proteins can be used for animal rations.

The introduction of ricin into heterologous systems is also being pursued for uses related to biodefense, treatment of AIDS, cancer immunotherapy, and disease model systems involving apoptosis. Production of pharmaceutically important compounds in plants is an area of interest that has expanded during the past two decades. However, the development of genetically engineered therapeutics utilizing type II RIPs such as ricin has been limited owing to the difficulty of their expression in heterologous systems due to improper polypeptide folding, incomplete processing of preproricin, and an overall susceptibility of host ribosomes to depurination (Frankel 1992). Successful processing of preproricin using either prokaryote or eukaryote expression systems has not been

reported, while Sehnke et al. (1994) demonstrated the processing of RIP preprotein precursor into a fully active mature toxin when expressed in tobacco. Sehnke and Ferl (1999) developed stable cell cultures from transgenic tobacco expressing ricin that provides a safe and simple means to produce properly processed recombinant ricin. Regardless of its advantages, the development and release of genetically modified ricin-producing species should be approached with caution, following appropriate safety regulations.

14.6.4 Modification of Oil Quality

The presence of the mid-chain hydroxyl group and double bonds imparts unique chemical and physical properties to castor oil that stabilizes the oil against oxidation, making castor oil a vital industrial feedstock. Castor oil contains ricinoleic acid (the 18-hydroxy oleic acid D-12-hydroxy octadec-cis-9-enoic acid), which is industrially important as it is used in the manufacture of polymers, lubricants, polyurethane coatings, cosmetics, and plastics. The ricinoleic acid molecule has three elements of functionality that make castor oil suitable for many chemical reactions and modifications (Ogunniyi 2006): (a) a carboxyl group that allows a wide range of esterifications, (b) a double bond that can be altered by hydrogenation, epoxidation, or vulcanization, and (c) a mid-chain hydroxyl group that can be acetylated, alkylated, or dehydrated to produce semidrying oils.

The uniqueness of castor bean not only lies in the presence of oleic acid hydroxylase to synthesize ricinoleic acid, but also in its capacity to efficiently accumulate high amounts of ricinoleic acid (88–91%) in its seed oil. Most of the genes encoding the key enzymes involved in the biosynthesis of fatty acids and triacylglycerols are single copy genes (Chan et al. 2010). Key areas of research to achieve desired oil quality include increasing oleic acid content in castor bean or producing ricinoleic acid in heterologous systems. Decreasing the ricinoleic acid content with concomitant increase in oleic acid content would enhance the performance of castor oil as a feedstock for biodiesel. Conventional breeding led to the identification of a castor bean mutant line (OLE-1), which has a 20-fold increase in oleic acid (C18:1) and a sixfold decrease in ricinoleic acid content (C18:1-OH) (Rojas-Barros et al. 2005). These studies concluded that the desaturation of oleic to linoleic acid was controlled by a major gene whose action was modified by a second gene. The recessive gene *ol* present in OLE-1 could affect the action of the oleoyl-12-hydroxylase enzyme preventing the hydroxylation of oleic acid to synthesize ricinoleic acid. The recessive alleles at the modifier (*M1*) locus would suppress the effect of the *ol* allele on the oleic/ricinoleic trait.

Genetic engineered production of ricinoleate in heterologous systems in order to replace castor bean has been a scientific and technical challenge. It would benefit ricinoleate-related industries by producing a valuable renewable resource in greater quantity without the complications of castor bean toxins. Incorporation, modification, or suppression of key metabolic pathways leading to over-accumulation of desired fatty acids are viable metabolic engineering approaches. Although it is relatively easy to genetically engineer biosynthetic pathways, achieving high levels of the industrial fatty acids in heterologous systems is not always successful. For example, transgenic expression of the fatty acid hydroxylase (*Fah12*) gene from castor bean under the control of strong seed-specific promoter (*napin*) from *Brassica napus* resulted in very low levels of accumulation of ricinoleic acid in the heterologous host (<1% in tobacco, 17% in Arabidopsis), making the approach commercially inviable (Brown and Somerville 1997). The major limitation for the accumulation of unusual fatty acids (UFAs) such as ricinoleate in heterologous systems is the inefficient transfer of the UFAs from the site of synthesis on phospholipids to the triacylglycerols (TAG) where the oil is assembled and stored. Lin et al. (2002) demonstrated that labeled ricinoleate is preferentially inserted into TAG in castor bean endosperm microsomes, while the same substrate failed to be efficiently incorporated in other

oilseeds. This is mainly because the castor bean acyl-CoA:diacylglycerol acyltransferase (DGAT1) selectively incorporates ricinoleate into TAG. The identification of the castor bean-specific ricinoleate gene phospholipid:diacylglycerol acyltransferase *PDAT1-2* and the manipulation of the phospholipid editing system in transgenic Arabidopsis plants enhanced the accumulation of total hydroxyl fatty acids (HFA) up to 25% (Kim et al. 2011). Subsequent studies involving incorporation of gene combinations, such as DGAT type 2 (*RcDGAT2*) and oleoyl hydroxylase (Burgal et al. 2008), or *PDAT* and oleoyl hydroxylase (van Erp et al. 2011), increased HFA content by 30% in transgenic plants. The conservation of seed oil metabolic pathways across plant species and knowledge of the mechanisms of seed oil accumulation has facilitated the development of designer oil crops by metabolic engineering of fatty acid profiles, and it can be applied to improve castor bean oil content and quality. The expression profiles of genes involved in fatty acid biosynthesis, transport during cellular endosperm development, and accumulation of key enzymes of the triacylglycerol biosynthetic pathway such as diacylglycerol acyltransferase, glycerol-3-phosphate dehydrogenase, and lysophosphatidyl-acyltransferases have been investigated in castor bean (He et al. 2004; Chen et al. 2007; Chan et al. 2010; Xian-Jie et al. 2011; Arroyo-Caro et al. 2013). Reverse engineering strategies are being developed wherein the TAG assembly routes are characterized as a prelude for transgenic expression of the specialized TAG assembly enzymes in oilseeds (Green et al. 2008). Under the Crop Biofactories Initiatives at CSIRO, Australia, several genes involved in the synthesis and storage of UFAs are being cloned and technologies are being developed to engineer high-level synthesis and accumulation of UFAs in transgenic oilseeds including safflower (Green et al. 2008). Expression of splice variants of the specific castor bean *wrinkled1* (*WRI1*) gene belonging to apetala 2 (*AP2*)/ethylene responsive element binding protein (*EREBP*) class of transcription factors in tobacco increased oil content by 4.3- to 4.9-fold as compared to controls and could be potentially used to engineer crops for high oil production (Ji et al. 2018).

Genetic engineering of ricinoleate synthesis has been attempted in *Lesquerella fendleri* for safe castor oil production. *L. fendleri* (L) is valued for its unusual hydroxy fatty acid called lesquerolic acid (C20:1OH), which is produced by a two-carbon elongation of ricinoleate. Therefore, suppression of the elongation step in *L. fendleri* through genetic engineering could generate ricinoleate in *L. fendleri* (Chen 2009).

Ectopic overexpression of castor bean leafy cotyledon (*LEC2*) in Arabidopsis triggered the expression of genes that encode regulators of seed maturation and oil body proteins in vegetative tissues. Consequently, expression of fatty acid elongase 1 (*FAE1*) was increased, inducing the accumulation of triacylglycerols, especially those containing the seed-specific fatty acid eicosenoic acid ($20:1^{\Delta 11}$) in vegetative tissues (Kim et al. 2014).

Fluctuating oil prices, depleting fossil fuel reserves, increasing concerns about climate change, and potential economic opportunity demand renewable sources of fuel. The feedstocks used for biodiesel include edible oil, but the debate on 'food versus fuel' indicates the need for introduction of alternative feedstocks which meet the criterion of (1) sustainability—not compete with food crops for land or water resources, not promote deforestation, the total life cycle of greenhouse gases emissions should be low compared to that of fossil fuels, (2) renewable and continuous availability—reliable supply with potential for large-scale production, and (3) economically feasible—competitive price compared to edible oils and fossil fuels.

Castor bean has attracted the attention of researchers in USA and Brazil because of its high oil content (50%) and relatively high crop yield, which can produce up to 140 gallons of castor oil per acre and provides a safe opportunity for biodiesel production without the risk of displacing food crops. Castor is a summer-grown crop that fits well in Mississippi and other parts of the Southern USA. Estimated CO_2 absorption level

of castor bean plants is 34.6 ton per ha with two growing cycles per year. Lifecycle analysis of biodiesel produced from castor bean showed that greenhouse gas emissions were reduced by 90% when compared to petroleum diesel. Castor oil is the only oil soluble in alcohol and therefore does not have the consequent energy requirement for transesterification as other vegetable oils. Castor oil has a good shelf life when compared to other oils (e.g., four times more stable than olive oil), and it does not turn rancid when subjected to excessive heat. However, its high viscosity, water content, and compressibility, as well as the observed ~10% reduction of its hydroxyl and acid values after 90 days of storage and its premium price, are the major issues limiting the use of straight castor oil as fuel for internal combustion engines (Scholz and da Silva 2008), although the limit values of viscosity, density, and cetane number can be met through transesterification followed by dilution or blending with conventional diesel fuel and other vegetable oils (www.castoroil.in).

Evogene (www.evogene.com) has demonstrated the potential of castor oil as a viable and sustainable second-generation feedstock for the production of biojet fuel as it meets the American Society of Testing and Materials (ASTM) D7566 fuel specification requirements for alternative aviation fuels containing synthetic hydrocarbons. Conversion of ricinoleic-rich castor oil to oleic-rich castor oil can be accomplished through silencing of the fatty acid hydroxylase (*fah1*) gene. Almeria-based researchers have genetically modified castor bean for obtaining oil with a higher proportion of monounsaturated fatty acids (e.g., oleic or palmitic) for use of castor oil as a biolubricant (https://cordis.europa.eu/news/rcn/119449_fr.html).

14.7 Conclusions and Future Perspectives

There is an immediate need for the development of a highly efficient, reliable, and reproducible direct and callus-mediated tissue culture system as a prelude for genetic engineering of castor bean for desirable traits. Transgenic approaches through RNAi technology has been proposed for reduction of the toxic protein ricin and conversion of ricinoleic acid-rich castor oil to oleic-rich oil. Bioengineering of fatty acid metabolic pathways in castor bean has a huge potential for downstream processing and value chain expansion. The availability of draft genome sequences of castor bean (Chan et al. 2010) and other Euphorbiaceae members (e.g., cassava and physic nut) allows genome-wide comparative analyses of stress-responsive genes, fatty acid metabolism genes, the ricin gene family. Their expression profiles help in the identification of key genes, promoters, and transcription factors as suitable targets for genetic engineering. The toxicity of castor bean seed cake due to ricin and RCA in animal feeding experiments poses biosafety and ethics concerns, and hence, suitable strategies should also be simultaneously developed for the assessment of the biosafety of castor transgenics.

With regard to biotic stresses, resistance to capsule borer and *Botrytis* gray rot could be incorporated through transgenic approaches. Considerable research gaps exist in our understanding of these pests' biology in different hosts, their behavior, population dynamics, seasonal abundance, off-season survival, host preference, pest–parasitoid relationships, economic thresholds, and management with suitable insecticides for effective control. The efficacy of crystal protein genes from *Bt* and other proteins against capsule borer has to be evaluated, and the rearing technique is an essential prerequisite for this purpose. Suitable candidate genes with appropriate deployment strategies need to be developed for enhancing resistance to *Botrytis* gray rot.

References

Ahn YJ, Chen GQ (2008) *In vitro* regeneration of castor *(Ricinus communis* L.) using cotyledon explants. HortScience 43:215–219

Ahn YJ, Vang L, McKeon TA, Chen GQ (2007) High-frequency plant regeneration through adventitious shoot formation in castor (*Ricinus communis* L.). Vitro Cell Dev Biol Plant 43:9–15

Alam I, Sharmin SM, Mondal SC, Alam MJ, Khalekuzzaman M et al (2010) *In vitro* micropropagation through cotyledonary node culture of castor bean (*Ricinus communis* L.). Aust J Crop Sci 4:81–84

Ankineedu G, Sharma KD, Kulkarni LG (1968) Effect of fast-neutrons and gamma rays on castor. Indian J Genet 28:31–39

Arroyo-Caro JM, Chileh T, Kazachkov M, Zou J, Alonso DL, Garcia-Maroto F (2013) The multigene family of lysophosphatidate acyltransferase (LPAT)-related enzymes in *Ricinus communis*: cloning and molecular characterization of two LPAT genes that are expressed in castor seeds. Plant Sci 200:29–40

Athma P, Reddy TP (1983) Efficiency of callus initiation and direct regeneration from different explants of castor (*Ricinus communis* L.). Curr Sci 52:256–257

Auld DL, Rolfe RD, McKeon TA (2001) Development of castor with reduced toxicity. J New Seeds 3:61–69

Auld DL, Pinkerton SD, Boroda E, Lombard KA, Murphy CK et al (2003) Registration of TTU-LRC castor germplasm with reduced levels of ricin and RCA_{120}. Crop Sci 43:746–747

Bahadur B, Reddy KRK, Rao GP (1992) Regeneration potential of callus cultures in castor (*Ricinus communis* L.). Asian J Plant Sci 4:13–18

Baldanzi M, Pugliesi C (1998) Selection for non-branching in castor, *Ricinus communis* L. Plant Breed 117:392–394

Birch RG (1997) Plant transformation: problems and strategies for practical application. Annu Rev Plant Physiol Plant Mol Biol 48:297–326

Brigham RD (1980) Castor. In: Fehr WR, Hadley HH (eds) Hybridization of crop plants. ASA, CSSA, Madison, WI, pp 235–247

Brown P, Somerville C (1997) Accumulation of ricinoleic, lesquerolic and densipolic acids in seeds of transgenic arabidopsis plants that express a fatty acyl hydroxylase cDNA from castor bean. Plant Physiol 113:933–942

Burgal J, Shockey J, Lu CF, Dyer J, Larson T et al (2008) Metabolic engineering of hydroxy fatty acid production in plants: RcDGAT2 drives dramatic increases in ricinoleate levels in seed oil. Plant Biotechnol J 6:819–831

Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J et al (2010) Draft genome sequence of the oilseed species *Ricinus Communis*. Nat Biotechnol 28:951–956. https://doi.org/10.1038/nbt.1674

Chen GQ (2009) Engineering *Lesquerella* for safe castor oil production. The association for the advancement of industrial crops (AAIC). 21st annual meeting in Termas de Chillan, Chillan, Chile, 14–19 Nov 2009

Chen GQ, Turner C, He X, Nguyen T, McKeon TA et al (2007) Expression profiles of genes involved in fatty acid and triacylglycerol synthesis in castor bean (*Ricinus communis* L.). Lipids 42:263–274

Chen Y, Wang Y, Wang Y, Huang F, Li G et al (2013) Construction of RNAi binary vector of ricin a chain and its transformation. Acta Bot Boreali-Occidentalia Sin 01

Duraimurugan P, Lakshminarayana M, Vimala Devi PS (2015) Comparative efficacy of microbial, botanical and chemical insecticides against lepidopteran pests in castor. Ecoscan 9:7–10

Frankel AE (ed) (1992) Genetically engineered toxins. Marcel Dekker, New York

Ganesh Kumari K (2010) *In vitro* culture and *Agrobacterium*-mediated genetic transformation studies in castor (*Ricinus communis* L. cv. TMV5) for improved tolerance to fungal diseases. Ph.D. thesis. Bharathidasan University, Tiruchirapalli, Tamil Nadu, India

Ganesh Kumari K, Ganesan M, Jayabalan N (2008) Somatic embryogenesis and plant regeneration in *Ricinus communis* L. Biol Plant 52:17–25

Gedil M, Kolade F, Raji A, Ingelbrecht I, Dixon A (2009) Development of molecular genomic tools for verification of intergeneric hybrids between castor bean (*Ricinus communis* L.) and cassava (*Manihot esculenta* Crantz). J Food Agric Environ 7:534–539

Gelvin SB (2003) *Agrobacterium*-mediated plant transformation: the biology behind the "gene-jockeying" tool. Microbiol Mol Biol Rev 67:16–37

Genyu Z (1988) Callus formation and plant regeneration from young stem segments of *Ricinus communis* L. Genetic Manipulation in Crops. IRRI, Cassell Tycooly, p 393

Green A, Singh S, Zhou XR (2008) Reverse engineering of novel industrial fatty acids in seed oils. In: Presented in the 7th International Safflower Conference, Wagga Wagga, Australia

Harlan JR (1976) Genetic resources in wild relatives of crops. Crop Sci 16:329–333

Hartley MR, Lord JM (1993) Structure, function and applications of ricin and related cytotoxic proteins. In: Grierson D (ed) Biosynthesis and manipulation of plant products, vol 3. Blackie Academic and Professional Publishers, London, pp 210–239

He X, Turner C, Chen GQ, Lin JT, McKeon TA (2004) Cloning and characterization of a cDNA encoding diacylglycerol acyltransferase from castor bean. Lipids 39:311–318

James C (2016) Global status of commercialized biotech/GM Crops: ISAAA *Brief* No. 52. ISAAA, Ithaca, NY

Ji XJ, Mao X, Hao QT, Liu BL, Xue JA et al (2018) Splice variants of the castor *WRI1* gene upregulates fatty acid and oil biosynthesis when expressed in tobacco leaves. Intl J Mol Sci 19:146

Kermode AR, Pramanik SK, Bewley JD (1989) The role of maturation drying in the transition from seed development to germination. VI. Desiccation-induced changes in messenger RNA populations within the endosperm of *Ricinus communis* L. seeds. J Exp Bot 40:33–41

Kim HU, Lee KR, Go YS, Jung JH, Suh MC et al (2011) Endoplasmic reticulum-located PDAT1-2 from castor bean enhances hydroxy fatty acid accumulation in transgenic plants. Plant Cell Physiol 52:983–993

Kim HU, Jung SJ, Lee KR, Kim EH, Lee SM et al (2014) Ectopic overexpression of castor bean LEAFY

COTYLEDON2 (LEC2) in Arabidopsis triggers the expression of genes that encode regulators of seed maturation and oil body proteins in vegetative tissues. FEBS Open Bio 4:25–32

Kolte SJ (1995) Insect pests of castor. In: Kolte SJ (ed) Castor diseases and crop improvement. Shipra Publications, New Delhi, India, p 95p

Komor E (1977) Sucrose uptake by cotyledons of *Ricinus communis* L. Characteristics, mechanism and regulation. Planta 137:119–131

Kulkarni LG, Ankineedu G (1966) Isolation of pistillate lines in castor for exploitation of hybrid vigour. Indian J Genet 26:363–365

Kumar MA, Rohini S, Kalpana N, Ganesh PT, Udayakumar M (2011) Amenability of castor to an *Agrobacterium*-mediated *in planta* transformation strategy using a *cry1AcF* gene for insect tolerance. J Crop Sci Biotechnol 14:125–132

Kumar SD, Tarakeswari M, Lakshminarayana M, Sujatha M (2016) Toxicity of *Bacillus thuringiensis* crystal proteins against eri silkworm, *Samia cynthia ricini* (Lepidoptera: Saturniidae). J Invertebr Pathol 138:116–119

Lakshminarayana M, Raoof MA (2005) Insect pests and diseases of castor and their management. Directorate of Oilseeds Research, Hyderabad, India, p 78

Lakshminarayana M, Sujatha M (2005) Toxicity of *Bacillus thuringiensis* var. *Kurstaki* strains and purified crystal proteins against *Spodoptera litura* (Fabr.) on castor, *Ricinus communis* (L.). J Oilseeds Res 22:433–434

Li W, Yong G, Wang Y, Wang CA (2014) *Agrobacterium*-mediated genetic transformation and regeneration of low-toxicity variety of castor (*Ricinus communis* L.). J Pure Appl Microbiol 8:1387–1395

Li W, Li Z, Zhai Y, Wang C (2015a) A highly efficient castor regeneration system identified through *WUSCHEL* expression. Chem Eng Trans 46:1393–1398

Li W, Zhai Y, Yang G, Li Z, Wang C (2015a) Optimization of regeneration technology on an energy crop-castor (*Ricinus communis* L.). In: International Power Electronics and Materials Engineering Conference (IPEMEC 2015), Atlantis Press, pp 1024–1026

Lin JT, Chen JM, Liao LP, McKeon TA (2002) Molecular species of acylglycerols incorporating radiolabeled fatty acids from castor (*Ricinus communis* L.) microsomal incubations. J Agric Food Chem 50:5077–5081

Lippincott JA, Haberlin GT (1965) The induction of leaf tumours by *Agrobacterium tumefaciens*. Amer J Bot 52:396–403

Lord JM, Lamb FI, Roberts LM (1984) Ricin: structure, biological activity and synthesis. Oxford Surveys Plant Mol Cell Biol 1:85–101

Malathi B, Ramesh S, Rao KV, Reddy VD (2006) *Agrobacterium*-mediated genetic transformation and production of semilooper resistant transgenic castor (*Ricinus communis* L.). Euphytica 147:441–449

Malik VS (2013) Use of genome editing technologies for improving castor bean and guar. J Plant Biochem Biotechnol 22:357–358

McKeon TA, Chen GQ (2003) Transformation of *Ricinus communis*, the castor plant. US Patent No 6.620.986

Molina SM, Schobert C (1995) Micropropagation of *Ricinus communis* L. J Plant Physiol 147:270–272

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant 15:473–497

Ogunniyi DS (2006) Castor oil: a vital industrial raw material. Bioresour Technol 97:1086–1091

Orlich G, Komor E (1992) Phloem loading in *Ricinus* cotyledon sucrose pathways *via* the mesophyll and the apoplasm. Planta 187:460–474

Patel MK, Joshi M, Mishra A, Jha B (2015) Ectopic expression of *SbNHX1* gene in transgenic castor (*Ricinus communis* L.) enhances salt stress by modulating physiological process. Plant Cell Tiss Org Cult 122:477–490

Peles Y, Lemer A, Meiri A, Dodek I, Gelfman S et al (2017) Determinate castor. Patent No WO 2017009847 A1

Pinkerton SD, Rolfe RD, Auld DL, Ghetie V, Lauterbach BF (1999) Selection of castor with divergent concentrations of ricin and *Ricinus communis* agglutinin. Crop Sci 39:353–357

Reddy KRK, Bahadur B (1989) *In vitro* multiplication of castor. In: Farook SA, Khan IA (eds) Recent advances in genetics and cytogenetics. Premier Publishers, Hyderabad, India, pp 479–482

Reddy KRK, Ramaswamy N, Bahadur B (1987a) Cross incompatibility between Ricinus and Jatropha. Plant Cell Incomp Newslett 19:60–65

Reddy KRK, Rao GP, Bahadur B (1987b) *In vitro* morphogenesis from seedling explants and callus cultures of castor (*Ricinus communis* L.). Phytomorphology 37:337–340

Richharia RH (1937) A note on the cytogenetics of *Ricinus communis* L. Indian J Agric Sci 7:707–711

Robinson SP, Beevers H (1981) Amino acid transport in germinating castor bean seedlings. Plant Physiol 68:560–566

Rojas-Barros P, de Haro A, Fernandez-Martinez JM (2005) Inheritance of high oleic/low ricinoleic acid content in the seed oil of castor mutant OLE-1. Crop Sci 45:157–162

Sailaja M, Tarakeswari M, Sujatha M (2008) Stable genetic transformation of castor (*Ricinus communis* L.) via particle gun-mediated gene transfer using embryo axes from mature seeds. Plant Cell Rep 27:1509–1519

Sangduen N, Pongtongkam P, Ratisoontorn P, Jampatas R, Suputtitada S et al (1987) Tissue culture and plant regeneration of castor (*Ricinus communis* L.). SABRAO J Breed Genet 19:144

Sarvesh A, Ram Rao DM, Reddy TP (1992) Callus initiation and plantlet regeneration from epicotyl and

cotyledonary explants of castor (*Ricinus communis* L.). Adv Plant Sci 5:124–128

Schobert C, Komor E (1992) Transport of nitrate and ammonia into the phloem and the xylem of *Ricinus communis* seedlings. Plant Physiol 140:306–309

Scholz V, da Silva JN (2008) Prospects and risks of the use of castor oil as a fuel. Biomass Bioenerg 32:95–100

Sehnke PC, Ferl RJ (1999) Processing of preproricin in transgenic tobacco. Protein Expr Purif 15:188–195

Sehnke PC, Pedrosa L, Paul AL, Frankel AE, Ferl RJ (1994) Identification of vacuolar expression of active, processed ricin in transgenic tobacco. J Biol Chem 269:22473–22476

Severino LS, Cordoba-Gaona OJ, Zanotto MD, Auld DL (2012) The influence of the caruncle on the germination of castor seed under high salinity or low soil water content. Seed Sci Technol 40:139–143

Sharma HC, Sharma KK, Seetharama N, Ortiz R (2000) Prospects for using transgenic resistance to insects in crop improvement. Mol Biol Genet 3:1–28

Singh D (1976) Castor-*Ricinus communis* L. (Euphorbiaceae). In: Simmonds NW (ed) Evolution of crop plants. Longman, London, pp 84–86

Singh PC, Kumar M, Chaturvedi CP, Yadav D, Tuli R (2004) Development of a hybrid δ-endotoxin and its expression in tobacco and cotton for control of a polyphagous pest, *Spodoptera litura*. Transgen Res 13:397–410

Sousa NL, Cabral GB, Vieira PM, Baldoni AB, Francisco JL et al (2017) Bio-detoxification of ricin in castor bean (*Ricinus communis* L.) seeds. Sci Rep 7:15385

Sujatha M (1996) Genetic and tissue culture studies in castor (*Ricinus communis* L.) and related genera. Ph. D. thesis, Osmania University, Hyderabad, India

Sujatha M, Lakshminarayana M (2005) Susceptibility of castor semilooper, *Achaea janata* L. to insecticide crystal proteins from *Bacillus thuringiensis*. Indian J Plant Protec 33:286–287

Sujatha M, Reddy TP (1998) Differential cytokinin effects on the stimulation of *in vitro* shoot proliferation from meristematic explants of castor (*Ricinus communis* L.). Plant Cell Rep 17:561–566

Sujatha M, Reddy TP (2007) Promotive effect of lysine monohydrochloride on morphogenesis in cultured seedling and mature plant tissues of castor (*Ricinus communis* L.). Indian J Crop Sci 2:279–286

Sujatha M, Sailaja M (2005) Stable genetic transformation of castor (*Ricinus communis* L.) via *Agrobacterium tumefaciens*-mediated gene transfer using embryo axes from mature seeds. Plant Cell Rep 23:803–810

Sujatha M, Reddy TP, Mahasi MJ (2008) Role of biotechnological interventions in the improvement of castor (*Ricinus communis* L.) and *Jatropha curcas* L. Biotechnol Adv 26:424–435

Sujatha M, Lakshminarayana M, Tarakeswari M, Singh PK, Rakesh Tuli R (2009) Expression of the *cry1EC* gene in castor (*Ricinus communis* L.) confers field resistance to tobacco caterpillar (*Spodoptera litura* Fabr) and castor semilooper (*Achoea janata* L.). Plant Cell Rep 28:935–946

van Erp H, Bates PD, Burgal J, Shockey J, Browse J (2011) Castor phospholipid:diacylglycerol acyltransferase facilitates efficient metabolism of hydroxy fatty acids in transgenic Arabidopsis. Plant Physiol 155:683–693

Weig A, Komor E (1992) The lipid transfer protein C of *R. communis* L. Isolation of two cDNA sequences which are strongly and exclusively expressed in cotyledons after germination. Planta 187:367–371

Weiss EA (1983) Castor. In: Weiss EA (ed) Oilseed crops. Longman, London, pp 31–99

Xian-jie M, Rong-hua X, Aizhong L, Ding W, Bo T (2011) Cloning and characterization of glycerol-3-phosphate dehydrogenase gene (RcGPDH) from castor bean. Chin J Oil Crop Sci 33:451–458

Zalavadiya VK, Mehta DR, Javia RM, Padhiyar SM, Madariya RB (2014) Somatic organogenesis and plant regeneration in castor (*Ricinus communis* L.). Asian J Bio Sci 9:43–52

Zhang J, Wang X, Feng Z, Geng X, Mu S et al (2016) *In vitro* establishment of a highly effective method of castor bean (*Ricinus communis* L.) regeneration using shoot explants. J Integr Agri 15:1417–1422