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.