Govind Singh Saharan
Naresh K. Mehta
Prabhu Dayal Meena

Genomics of Crucifer's Host-Pathosystem

Springer

Genomics of Crucifer's Host-Pathosystem

Govind Singh Saharan • Naresh K. Mehta •
Prabhu Dayal Meena

Genomics of Crucifer's Host-Pathosystem

Springer

Govind Singh Saharan
Department of Plant Pathology
CCS Haryana Agricultural University
Hisar, Haryana, India

Naresh K. Mehta
Department of Plant Pathology
CCS Haryana Agricultural University
Hisar, Haryana, India

Prabhu Dayal Meena
Plant Protection Unit
ICAR-Directorate of Rapeseed-Mustard Research
Bharatpur, Rajasthan, India

ISBN 978-981-19-3811-5 ISBN 978-981-19-3812-2 (eBook)
https://doi.org/10.1007/978-981-19-3812-2

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Foreword

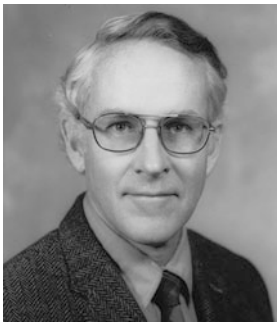
The family Brassicaceae include 338 genera with 3709 species of which the genus *Brassica* comprises 37 species. A number of *Brassica* species are of global economic importance for yielding quality edible oil, nutritious vegetables, and animal forage and as with most crops, Brassicas are subject to considerable losses caused by biotrophic, necrotrophic, and hemibiotrophic pathogens. Recent advances in molecular techniques have greatly expanded our understanding of plant pathology, disease resistance, and immunity, enabling their application through plant breeding to agricultural crop systems. Within the Brassicas, host-pathosystems of various crops are being studied genomically using multiomic research technologies.

Plant immune/defense systems rely on their ability to recognize the presence of the pathogen, to carry out signal transduction, and to respond defensively through pathways involving many genes and/or their products. The plant immune systems have provided both increased evolutionary opportunity for pathogen resistance and additional mechanisms for pathogen inhibition of such defense responses. However, sequencing the genomes of plant pathogens has shown variability in genome size and its structure. Genomic data generated so far has provided insightful information to understand the mechanisms and the changes in genome size and adaptive evolution in plant pathogens. The pathogens provide prominent advantageous associations between rapidly evolving transposable elements and virulence genes that cause variation in virulence of the phenotypes. The genome architecture, pathogenicity factors, and determinants of host specificity of some pathogens are yet to be known.

The outstanding book *Genomics of Crucifer's Host-Pathosystem* authored by Prof. (Dr.) G. S. Saharan, Prof. (Dr.) Naresh K. Mehta, and Dr. Prabhu Dayal Meena is the ninth book in the series on diseases of crucifers being published by Springer, Nature, after the books on *Albugo*, *Alternaria*, *Erysiphe*, *Hyaloperonospora*, *Plasmodiophora*, *Sclerotinia*, Crucifer's host resistance, and *Molecular mechanism of crucifer's host resistance*. The book is an annotated compilation that explains the characteristics of the recent development in genome assembly of important pathogens of crucifers to gain insights into host-pathosystem. The manuscript has been prepared after critical analysis of the world literature on all the major diseases of crucifers with encyclopedic treatments for better comprehension by readers. The

book provides helpful background and current information that enhances our insight on host-pathosystem of crucifers.

My heartiest congratulations to the authors for bringing out their lifelong professional interest, and expertise with comprehensive treatise at the most needed time. It will be quite useful for researchers, teachers, students, extension workers, and all those who are concerned with the growth and development of cruciferous crops.



Department of Plant Pathology,
University of Wisconsin
Madison, WI, USA

Paul H. Williams

Preface

Crucifer encompasses one of the major groups of crops belonging to the family Brassicaceae yielding high-quality edible oil and nutritious vegetables. Major diseases of these crops cause heavy yield and quality losses at global level which are increasing with the climate change threatening food security. The use of omics technologies allows us to understand the genomics of crucifers' host-pathosystem at molecular level to get deeper insights into the interacting host and pathogen to reveal systems biology. The integrated multiomics and systems biology data will enable breeding of high-quality disease-resistant *Brassica* crops in a more holistic, targeted, and precise way for better management of long-term co-evolutionary process of host and pathogen and in desired direction. Application of omics approaches has allowed getting high-quality genome assemblies of all the six *Brassica* species of "U" triangle in addition to the model plant *Arabidopsis thaliana* and major pathogens of *Brassica* crops. It has greatly facilitated the discovery of candidate gene for effectors and virulence factors of the pathogens. Transcriptomics of virulence-related genes has been applied to study pathogen gene expression during host invasion in order to monitor molecular mechanisms and events involved in pathogenesis. Secretomics revealed effector proteins and specific enzymes crucial for host colonization by the pathogens. Interactomics studied networks of genes and proteins interactions in biological host-pathosystem. Biometabolomics provided an insight into metabolic changes in the host during host-pathogen interactions and its effects on host-pathosystem.

The book *Genomics of Crucifer's Host-Pathosystem* is a compilation of latest research achievements of *Brassica* scientists using omics approaches in understanding host-pathogen interaction, molecular detection, identification, and functional characterization of effectors/genes including pathogenomics and biometabolomics, genomics of pathogenic variability, and genomics modulation of crucifers' host-pathosystem through nine chapters on all the important pathogens of crucifers. The book is a source of information which has been vividly illustrated with photographs, graphs, figures, histogram, tables, and colored plates, which makes it stimulating, effective, and easy to comprehend by readers. The headings and subheadings of each chapter have been arranged in numbered series to make the subject matter contiguous. A chapter on standardized, reproducible protocols in studying the host-pathosystem has been included for the researchers of cruciferous crops for

developing resistant cultivars after understanding the host-pathosystem. The last section deals with the gaps in understanding, knowledge of genomics, and offers suggestions for future research priorities in order to initiate advanced research programs in conducting studies to foster success in *Brassica* production and productivity through the development of improved disease-resistant varieties for the benefit of farmers.

The authors are confident that this book will be immensely useful to researchers especially *Brassica* breeders, molecular biologists, plant pathologists, teachers, extension specialists, students, industrialists, farmers, and all others who are interested to grow healthy and profitable cruciferous crops all over the world. Any shortcomings, lacunae, and flaws in the book are the responsibility of the authors. Any suggestions by readers are always a source of inspiration for the authors, and suggestions for its improvement are most welcome.

Hisar, Haryana, India — Govind Singh Saharan
Hisar, Haryana, India — Naresh K. Mehta
Bharatpur, Rajasthan, India — Prabhu Dayal Meena

Acknowledgments

The authors are highly grateful to all the persons, scientists, publishers, societies, journals, institutes, websites, and all others whose valuable materials such as photographs (macroscopic, microscopic, electron micrographs, scanning electron micrographs), drawings, figures, histograms, graphs, tables, and flowcharts, etc. have been used through reproduction in this book. The address of the author(s)/source(s) from where material was adopted/produced from institutes, websites, societies, etc. can be obtained through the reference which has been cited in the reference section of the book in each chapter. We, the authors, duly thankfully acknowledge and put on record all the authors whose excellent and valuable publications/articles find their place in this manuscript by citing their names against tables, figures, plates, dendrograms, etc. We have sought the necessary permission to reproduce their material in this book, and majority of them have granted the permission to reuse them. The authors are sincerely thankful to all the scientists, publishers, journals, institutes, societies, and websites whose materials have been reproduced in one or the other form in the present manuscript but forgot to acknowledge their name(s) inadvertently. Their prompt help in providing permission, information, photographs, figures, and consents for reproduction in the present manuscript has been duly acknowledged, and the authors are grateful to all of them. There might be some errors, mistakes, and shortcomings in this manuscript. We shall greatly appreciate healthy criticism and suggestions for the improvement of this publication in future.

Contents

1 Genomics of Crucifer's Host-Pathosystem: Prologue 1

2 Crucifer's Pathogens Genome 127

3 Genomics of Host–Pathogen Interaction 297

4 Molecular Detection and Identification of Pathogens, Pathotypes, and Genes 503

5 Pathogenomics of Pathogenic Variability 595

6 Biometabolomics of Crucifer's Host-Pathosystem 729

7 Genomics of Crucifer's Host-Pathosystem at a Glance 799

8 Protocols to Study Host-Pathosystems 831

9 Future Research Priorities of Crucifers' Host-Pathosystem 915

Index 923

About the Authors

Govind Singh Saharan, Ex. Professor and Head, Department of Plant Pathology, CCS HAU, Hisar, has conducted research in diverse fields of Plant Pathology and has published 250 articles in national and international journals. He has been editor/author of several books, monographs, and Crop Production Compendium (CABI). He is on the panel of Experts of SAU, ICAR, IARI, CSIR, UGC, and Department of Biotechnology. Dr. Saharan has been a visiting Professor at the University of Alberta, Edmonton, Canada; Agriculture and Agri-Food Canada, Saskatoon, Canada; and Rothamsted Research, IACR, Harpenden, UK. He has been President (NZ) of the Indian Phytopathological Society, Editor-in-Chief of a few journals, and President of the Indian Society of Mycology and Plant Pathology. He has been awarded with the Y. L. Nene Outstanding Plant Pathology Teacher Award by ISMPP, Udaipur, and Life Time Achievement Award by the Society for Rapeseed-Mustard Research, Bharatpur, India.

Naresh K. Mehta, Former Associate Dean, Professor and Consultant Faculty, Department of Plant Pathology, CCS HAU, Hisar, has been teaching Plant Pathology courses to UG and PG students. He has conducted research in diverse fields of Plant Pathology especially on rapeseed-mustard. He has guided several M.Sc. and Ph.D. students. He has published about 200 research articles and editor/author of several books, book chapters, review articles, and teaching manuals. Dr. Mehta has been President of INSOPP, Ludhiana, and Editor-in-Chief of a journal and served as a member of the editorial board of various phytopathological societies in India. Dr. Mehta has been awarded the

Dr. Y. L. Nene Outstanding Plant Pathology Teacher Award by ISMPP, Udaipur; Dr. T. S. Thind Distinguish Plant Pathologist Award by INSOPP, Ludhiana, and Ms. Manju Utereja Memorial Gold Medal, HAU, Hisar. He is on the panel of Experts of SAU, ICAR, UGC, and IARI and a member of various national and international committees. Dr. Mehta has been a visiting scientist to the University of Alberta, Edmonton, Canada.

Prabhu Dayal Meena, Principal Scientist (Plant Pathology) at ICAR-Directorate of Rapeseed-Mustard Research, Bharatpur, Rajasthan, India, has conducted research on different aspects of rapeseed-mustard diseases, including host resistance, management, epidemiology, and biology of crucifer's pathogens. He has associated in the development of six high yielding varieties and registered seven trait-specific genotypes. He has published 120 research papers, reviews, and book chapters and 13 books. Dr. Meena is Fellow of Indian Society of Mycology and Plant Pathology, Plant Protection Association of India, Indian Society of Oilseed Research, Society for Rapeseed-Mustard Research, and Indian Phytopathological Society. He has been honored with the Dr. P.R. Kumar Outstanding Brassica Scientist Award and Gold Medal of Society for Rapeseed-Mustard Research. He is founder secretary, managing editor, and Chief Editor of Society for Rapeseed-Mustard Research. He has visited the UK under Indo-UK Collaborative Research on Oilseed Brassica crops. He has been guiding several M.Sc. and Ph.D. students.

Abbreviations

AAFC	Agriculture and Agri-Food Canada
AB	Alternaria blight
ABA	Abscisic acid
ABC	ATP-binding cassette
ABP1	Auxin binding protein 1
AFLP	Amplified fragment length polymorphism
AG	Anastomosis groups
AIP	Aminoindan phosphonic acid
AM	Arbuscular mycorrhizal
ANOVA	*Analysis of variance*
AOC3	*Allene oxide cyclase 3*
AOS	Allene oxide synthase
ap	Appressorium
APX	Ascorbate peroxidase
ARF-GEF	ADP ribosylation factor-GTP exchange factor
AT	Associative transcriptomics
AtSTP4	*Arabidopsis sugar transport protein 4*
AUDPC	Area Under Disease Progress Curve
Avr	Avirulence
BA	Benzoic acid
BABA	β-amino butyric acid
BC	Backcross
Bgt	*Blumeria graminis* f. sp. *tritici*
BiFC	Bimolecular fluorescence complement
BLAST	Basic Local Alignment Search Tool
BLAT	BLAST-like alignment tool
BLUPs	Best linear unbiased predictions
BSA	Bulked Segregant Analysis
BSMV	Barley stripe mosaic virus
CA	Constitutively activated
CAC	Clathrin adaptor complex
CAD	Cinnamyl alcohol dehydrogenase

CaM	Calmodulin
CaMV	Cauliflower mosaic virus
CAT	Catalase
CBD	Chitin binding domain
CBls	Calcineurin-B-like protein
CC	Coiled coil
CDPKs	Calcium-dependent protein kinases
CF	Culture/cultural filtrate
ChIP	Chromatin immune-precipitation
CHS	Chalcone synthase
CI	Cylindrical inclusion protein
CK	Cytokinins
cM	Centi Morgans
CMLs	Calmodulin-like protein
CNGCs	Cyclic nucleotide gated ion channels
COIP	Co-immune precipitation
CW	Cell wall
CWAs	Cell wall apposition
CWDE	Cell wall-degrading enzymes
DAI	Days after inoculation
DAMPs	Damage-associated molecular patterns
DAPC	Discriminant analysis of principal components
DAS	Days after sowing
DEGs	Differentially expressed genes
DGE	Differential gene expression
DH	Doubled haploid
DI	Disease indices
DM	Downy mildew
DNA	Deoxyribonucleic acid
dpi	Days post-inoculation
DTT	Dithiothreitol
Ec	*Erysiphe cruciferarum*
ECD	European clubroot differential
EDS1	Enhanced disease susceptibility 1
EHM	Extrahaustorial membrane
ELISAs	Enzyme-linked immunosorbent assays
EMS	Ethyl methanesulfonate
ENU	Ethyl nitrosourea
Ep	*Erysiphe pisi*
e-PCR	Electronic PCR
ERK	Extracellular signal-regulated kinase
ESTs	Expressed sequence tags
ET	Ethylene
ETI	Effector-triggered immunity
ETS	Effector-triggered susceptibility

f. sp.	Fungal forma specialis
FCF	Fungal culture filtrate
FDR	False discovery rate
Fig	Figure
FPKM	Fragments Per Kilobase of transcript per Million mapped
FTIR	Fourier-transform infrared spectroscopy
GA	Gibberellins
GBS	Genotyping by sequencing
GC	*Golovinomyces cichoracearum*
GC-MS	Gas chromatography-mass spectrometry
GCRMA	Guanine cytosine robust multi-array analysis
GDIs	Guanine nucleotide dissociation inhibitors
GDP	Guaracol-dependent peroxidase
GFP	Green fluorescent protein
GISH	Genomic in situ hybridization
GLM	Generalized linear models
GM	Genetically modified
GO	Gene ontology
Go	*Golovinomyces orontii*
GPCR	G-protein-coupled receptors
GSS	Genomic survey sequences
GWAS	Genome-wide association analysis
HGP	Human Genome Project
HIGS	Host-induced gene silencing
HMM	Hidden Markov Model
HNRT	Homoeologous nonreciprocal transposition
Hp	*Hyaloperonospora parasitica*
Hpa	*Hyaloperonospora arabidopsidis*
hpi	Hours post inoculation
HR	Hypersensitive response
HRMS	High-resolution mass spectrometry
HS	Highly susceptible
HSPs	Heat shock proteins
HTGs	High-throughput genome sequences
HTS	Host targeting signal
IAA	Indole-3-acetic acid
IAN	Indole-3-acetonitrile
IAOx	Indole-3-acetaldoxime
IC	Isochorismate
ICAR	Indian Council of Agricultural Research
ICIM	Inclusive composite interval mapping
ICM	Composite interval mapping
ICS	Isochorismate synthases
iGS	Indole glucosinolates
ILs	Introgression lines

IM	Interval mapping
INA	Isonicotinic acid
IP	Intron polymorphic
IPG	Immobilized pH gradient strips
ISSRs	Inter-simple sequence repeats
ITCs	Isothiocyanates
ITS	Internal transcribed spacer
JA	Jasmonic acid
KASP	Kompetitive allele-specific PCR
LD	Linkage disequilibrium
LGs	Linkage groups
LIF	Lignification inducing factor
Lm	*Leptosphaeria maculans*
LOD	Logarithm of odds difference
LRR	Leucine-rich repeat
LRR-RLKs	Leucine-rich repeat receptor like kinase
LRR-RLPs	Leucine-rich repeat receptors–like protein
LRRs	Leucine-rich repeats
LysM	Lysine motif
LZ	Leucine zipper
MAB	Marker-assisted backcross breeding
MAMPs	Microbe-associated molecular patterns
MAP	Mitogen-activated protein
MAPK	Mitogen-activated protein kinase
MAPKKK	MAP kinase kinase kinase
MAS	Marker-assisted selection
MeSA	Methyl salicylate
MET	Multi-environment trials
Mips	Macrophage infectivity potentiators
MKK	MAP kinase kinase
ML	Maximum likelihood
MLM	Mixed linear model
MLO	Mildew resistance locus O
MPK	MAP kinase
MPTO	Methylthiopentanaldoxime
MQM	Multiple QTL mapping
MR	Moderately resistant
MS	Mass spectrometry
MSA	Multiple sequence alignment
MT	Microtubule
MTAs	Marker trait associations
MYA	Million years ago
MYB	Myeloblastosis
NBS	Nucleotide binding site
NBS-LRR	Nucleotide-binding site leucine-rich repeat

NBT	Nitro blue tetrazolium
NC	Nucleotidyl cyclase
NCBI	National Center for Biotechnology Information
NGS	Next generation sequencing
NHR	Nonhost resistance
NIa	Nuclear inclusion protein a
NILs	Near isogenic lines
NIRS	Near-infrared reflectance spectroscopy
NLRs	Nucleotide-binding site leucine-rich repeats
NMR	Nuclear magnetic resonance
NO	Nitric oxide
NPR1	Nonexpressor of pathogenesis-related genes 1
NWCVT	National winter canola variety trials
OA	Oxalic acid
OD	Optical density
OGs	Oligogalacturonides
On	*Oidium neolycopersici*
OST1	Open stomata 1
PA	Phosphatidic acid
PAD4	Phytoalexin-deficient 4
PAL	Phenylalanine ammonia lyase
PAMP	Pathogen-associated molecular pattern
PCA	Principal component analysis
PCD	Programmed cell death
PCR	Polymerase chain reaction
PDA	Potato dextrose agar
PDI	Percent disease index
PDR	Pleiotropic drug resistance
PEIs	Pectinesterase inhibitors
PEN2	Peroxisome-associated myrosinase penetration 2
PFK	Phosphofructokinase
PGA	Polygalacturonase
PGIPs	Polygalacturonase inhibitor proteins
PIC	Polymorphic information content
PIPs	Phosphatidyl inositol phosphates
PM	Plasma membrane
PM	Powdery mildew
PMEI	Pectin methylesterase inhibitors
PMSF	Phenylmethylsulphonyl fluoride
PO	Peroxidase
pp	Penetration peg formation
PPO	Polyphenol oxidase
PPT	Phosphinothricin
PR	Pathogenesis related
PRRs	Pattern recognition receptors

PTI	PAMPs-triggered immunity
PTI	Pattern-triggered immunity
pv.	Pathovar
QDR	Quantitative disease resistance
qRT-PCR	Real-time quantitative PCR
QTL	Quantitative trait loci
R	Resistance
RAPD	Random amplification of polymorphic DNA
RFLP	Restriction fragment length polymorphism
RGL	Resistant genes like
RIN	RNA integrity numbers
RLCK	Receptor-like cytoplasmic kinase
RLKs	Receptor-like kinases
RLPs	Receptor-like proteins
RNA	Ribonucleic acid
RNAi	RNA interference
ROI	Reactive oxygen intermediates
ROP	Rho of plants
ROS	Reactive oxygen species
RT-PCR	Reverse transcription and quantitative reverse transcription-polymerase chain reaction
SA	Salicylic acid
SAG	Salicylic acid glycoside
SAG101	Senescence-associated gene 101
SAM	*S*-adenosine-L-methionine
SAR	Systemic acquired resistance
SARF	Sum of adjacent recombination fractions
SCAR	Sequence characterized amplified region
SD	Standard deviation
Si	Silicon
SMA	Single marker analysis
SMRT	Single-molecule real-time
SNAP	Soluble *N*-ethylmaleimide-sensitive factor adaptor protein
SNP	Single nucleotide polymorphism
SOD	Superoxide dismutase
SRAP	Sequence-related amplified polymorphism
SSIL	*Sclerotinia sclerotiorum integrin-like*
SSR	Sclerotinia stem rot
SSR	Simple sequence repeat
STK	Serine-threonine kinase
STS	Sequence tagged sites
TAIR	The *Arabidopsis* Information Resource
TBIAs	Tissue blot immunoassays
TBS	Tris-buffered saline
TDFs	Transcript-derived fragments

TFs	Transcription factors
TGS	TRIS-Glycine-SDS
TIGS	Transient-induced gene silencing
TIR	Toll/Interleukin-1 receptor
TM	Trans-membrane
TMM	Trimmed mean of means
TNL	TIR-NBS-LRR
TuMV	Turnip mosaic virus
TuYV	Turnip yellows virus
UPGMA	Unweighted pair group method with arithmetic mean
UPS	Ubiquitin proteasome system
USA	United States of America
USDA	United States Department of Agriculture
UTR	Untranslated region
VAMPs	Vesicle-associated membrane proteins
VIGS	Virus-induced gene silencing
VPg	Genome-linked viral protein
WAKL	Wall-associated kinase-like
WGT	Whole genome triplication

Symbols

$	Dollar
%	Percent
≈	Is approximately equal to
/	Per
:	Ratio
~	Approximately
£	Pound
+	Plus
<	Less than
=	Is equal to
>	More than
±	Plus-minus sign
×	Multiply
≤	Less than or equal to
≥	More than or equal to
µg	Microgram
µl	Microliter
µm	Micrometer
µmol	Micromole
AU$	Australian dollar
B	Boron
bp	Base pair

Ca	Calcium
Ca^{2+}	Calcium cation
CFU	Colony forming unit
cM	Centi Morgans
cm	Centimeter
cm^2	Square centimeter
cm^{-3}	Per cubic centimeters
cvs.	Cultivars
d	Day
dpi	Days post-inoculation
dS/m	DeciSiemens per meter
e.g.	For example
et al.	"et alia" meaning "and others"
g	Gram
g^{-1}	Per gram
Gy	Gray
h	Hours
h^2	Heritability
H_2O_2	Hydrogen peroxide
i.e.	id est, meaning "that is"
kb	Kilobase
KCl	Potassium chloride
kV	Kilovolt
L	Liter
L^{-1}	Per liter
Log	Logarithm
M	Million
m	Meter
M	molar
m^2	Square meter
Mb	Million bases
Mb	Megabyte
Mg	Magnesium
mg	Milligram
$MgCl_2$	Magnesium chloride
Mha	Million hectare
min	Minute
ml	Milliliter
mL^{-1}	Per milliliter
mm	Millimeter
mM	Millimolar
mmole	Millimoles
mmt	Metric million tonnes
MW	Molecular weight
N	Nitrogen

nm	Nanometer
nmoles	Nanomole
OC	Degree Celsius
P	Potassium
pH	Potential of hydrogen
rpm	Revolutions per minute
S	Sulfur
S	Susceptible
s	Second
s^{-1}	Per second
sdH_2O	Sterile deionized water
SE	Standard error
Subsp.	Subspecies
V	Volts
v/v	Volume by volume
v:w	Volume equal to weight
var.	Variety
viz.	Videlicet meaning “namely” or “which is”
w/v	Weight by volume
β	Beta
μM	Micrometers
Σ	Sigma means “sum up”

Genomics of Crucifer's Host-Pathosystem: Prologue

1

Abstract

Crucifer's belong to the family Cruciferae containing over 3660 species which includes very important crops of human and animal needs yielding quality edible oil, industrial oilseeds, vegetables, and fodder crops. The largest group of crops contains 39 species of Brassicaceae grown all over the world including *Arabidopsis thaliana* which has been used widely for dissecting molecular mechanisms of crucifer's host-pathosystem through multiomics approaches. Under natural and cultivated conditions, crucifer's are challenged by several abiotic and biotic stresses viz., *Albugo* (White rust), *Alternaria* (Alternaria blight), *Colletotrichum* (Anthracnose), *Erysiphe* (Powdery mildew), *Fusarium* (Fusarium wilt), *Hyaloperonospora* (Downy mildew), *Leptosphaeria* (Blackleg), *Plasmodiophora* (Clubroot), *Pseudocercosporella* (White leaf spot), *Pyrenopeziza* (Light leaf spot), *Sclerotinia* (Stem rot), Turnip mosaic virus (*TuMV*), *Verticillium* (Verticillium wilt), *Xanthomonas* (Black rot), and *Heterodera* (Cyst nematode). These pathogens have been used as model host-pathosystem to reveal genomics of crucifer's host-pathosystem. The genomics of plants was initiated after the sequencing of *A. thaliana* genome for the first time in 1990. The genome size of *A. thaliana* is 125 Mbp containing 25,498 genes encoding proteins from 11,000 families. Now, the genome of all the six *Brassica* species has been sequenced. The sequence analysis has revealed genome size of *B. carinata* 642 Mb, *B. juncea* 922 Mb, *B. napus* 925 Mb, *B. nigra* 591 Mb, *B. oleracea* 584.60 Mb, and *B. rapa* 485 Mb. The genome of major pathogens of crucifer's has also been sequenced. The availability of genome sequence analysis of both host and pathogen has allowed rapid identification of candidate genes during different events of host–pathogen interaction to understand genes governing pathogenesis and host resistance. The application of omics technologies like NGS, Pangenomics, SNP, In Silico, BSA, Ren Seq, Effectoromics, Transcriptomics, Proteomics, Secretomics, Interactomics, and

G. Singh Saharan et al., *Genomics of Crucifer's Host-Pathosystem*,
https://doi.org/10.1007/978-981-19-3812-2_1

Metabolomics has benefitted greatly in revealing the complex biological, genetical, and molecular mechanisms of crucifer's host-pathosystem. Crucifers have developed molecular mechanisms in response to multiple stresses which are activated through complex signaling pathways for the expression of genes to overcome these stresses. Several genes are differentially expressed in response to multiple stresses. At global level, 16 pathogens causing crucifers diseases are considered as of major consequences based on their geographical distribution, host range, losses caused, and resources spend to manage them. Using crucifer's host-pathosystem with these pathogens, several genes have been identified, mapped on *Brassica* chromosomes with functional characterization, isolation, and cloning to produce resistant cultivars of *Brassica*. In recent years, QRT-PCR, a powerful and efficient technique has become the first choice in quantitative gene expression in crucifer's host-pathosystem for various biological and molecular functions. However, data normalization is essential for reliable output of QRT-PCR assays to avoid unsuitable choice of reference genes.

Keywords

Economic importance · Crucifers' · *Brassica* genomics · *Arabidopsis* · Omics approaches · Host-pathosystem · Secretomics · Metabolomics · Biotic and abiotic stress · Expression of genes · Major pathogens at global level · Their morphology · Disease symptomatology · Host range · Epidemiology · Fine structures · Pathogenic variability · Infection and pathogenesis · Host resistance · Disease management · Differential genes to overcome biotic stress · Signaling system of crucifers' · Genomics of crucifers host resistance

1.1 Introduction

The family Cruciferae is comprised of most diverse plants containing over 3660 species including important edible, industrial oilseeds, vegetable, and fodder crops. The largest group of crops belonging to family Brassicae contains 39 species, and includes six U's triangle crops grown all over the world for oilseeds, vegetables, and as forage crops. The *Arabidopsis thaliana* is also a member of this family which has become the most important crop to dissect molecular mechanisms and molecular genetical, interactions using *Arabidopsis* host-pathosystem as a model in the scientific community of the world.

Cruciferous crop plants like all other living organisms do not live in isolation but are under the influence of interactions among themselves as well as biotic (pathogen, pests, weeds), abiotic (temperature, salinity, alkalinity, drought), and environmental factors (above and below grounds) where they are cultivated or wild. Under natural conditions, they are always challenged simultaneously by several biotic, abiotic, and internal defense reactions against these stresses. To defend themselves against all kinds of stresses, plants employ pre- and post defense mechanisms triggered by signaling molecules. Multiple stress responses are coordinated by complex signaling

network molecules which converge at point of defense. The application of omics and genomics approaches has facilitated to have deeper insight into these complex interactions and use the modern genomics technology to breed stress resistant idiotype crucifers' crops. More than 50 pathogens are reported to cause diseases in crucifers' crops all over the world. Out of these, 16 pathogens are widely distributed causing heavy yield losses in cruciferous crops. The genomics of crucifer's host-pathosystem has been investigated at global level using these pathogens as model system. Each of the following factors has great significance and bearing on comprehension of this complex biological system and all events during each stage of interaction for better management of crucifers crops against all kinds of stresses with higher yield and quality of crops.

1.2 Economic Importance of Crucifers

Crucifers include large number of vegetables and oil yielding crops of family Cruciferae. The largest group of crops belongs to family Brassicae. The *Brassica* genus is a member of Brassicaceae (Cruciferae) and contains 39 species (http://www.theplantlist.org/). Among the *Brassica* species, six constitute U'sTriangle2: three diploid species, namely *Brassica rapa* (AA genome: $2n = 2x = 20$), *Brassica nigra* (BB: $2n = 2x = 16$), and *Brassica oleracea* (CC: $2n = 2x = 18$), and three allotetraploid species, namely *Brassica juncea* (AABB: $2x = 4x = 36$), *Brassica napus* (AACC: $2n = 4x = 38$), and *Brassica carinata* (BBCC: $2n = 4x = 34$). The triangle model provides the fundamental relationships among these *Brassica* species and is used as an important guideline for both evolutionary research and the improvement of *Brassica* crops via interspecies crossing to facilitate gene exchanges.

The crucifers' crops have great economic and morphological importance in daily life. Crops of the genus *Brassica* (tribe Brassiceae), which are in the same taxonomic family as *Arabidopsis thaliana*, are widely used in the cuisine of many cultures, due, in part, to the many choices of edible forms in the genus. Economically, *Brassica* is loosely categorized into oilseed, vegetable, and condiment crops. The *B. napus*, *B. rapa* (formerly *campestris*), *B. juncea*, and *B. carinata* provide about 12% of the world-wide edible vegetable oil supplies. The *B. oleracea* and *B. rapa*, the so-called "cole crops," comprise many of the vegetables in our daily diet. Several of these vegetables have extreme morphological characteristics. Examples of such morphologies include the enlarged inflorescence of cauliflower (*B. oleracea* subspecies *botrytis*) and broccoli (*B. oleracea* subspecies *italica*); the enlarged stem of kohlrabi (*B. oleracea* subsp. *gongylodes*) and marrow stem kale (*B. oleracea* subspecies *medullosa*); the enlarged root of turnip (*B. rapa* subspecies *rapifera*); the enlarged and twisted leaves of Pak-choi (*B. rapa* subspecies *chinensis*), Chinese cabbage (*B. rapa* subspecies *pekinensis*); and the enlarged single apical bud of cabbage (*B. oleracea* subspecies *capitata*) or the many auxiliary buds of Brussels sprout (*B. oleracea* subspecies *gemmifera*). Finally, the seed of *B. nigra* is utilized as a condiment-mustard. *Brassica* species are a valuable source of dietary fiber,

vitamin C, and other possible salubrious factors such as anticancer compounds. Estimates of the economic importance of *Brassica* species are conservative, because several cole crops such as collards are cultivated primarily for local or home use, but are none the less a dietary main stay in low-income communities where other fresh vegetables can be prohibitively expensive.

Crops belonging to the *Brassica* genus are among the ten most economically important vegetable crops in global agriculture and markets. Due to their wide adaptation and ability to thrive under varying agroclimatic conditions, *Brassica* crops are grown throughout the world for food, animal forage and fodder and also for industrial applications. They also have an allelopathic use for sustainable agriculture and are grown as phytore mediators against heavy metals such as cadmium. As far as food is concerned, now a day, consumers are demanding products that are rich in nutrients for optimal health benefits. In this respect, the popularity of *Brassica* products is increasing because of their nutritional value, and anticancer, antioxidant, and anti-inflammatory properties. Nutritionally, these vegetables are low-fat, have a high vitamin (C and E) content and contain minerals (P, S, Cl, Ca, Fe, Sr, K, Cr, Mn, Se, and Zn) and fiber. In addition, they contain important phytochemicals that are beneficial for human health, such as anthocyanins, flavonoids, terpenes, *S*-methyl cysteine sulfoxide, coumarins, and other small compounds. However, the most characteristic compounds are glucosinolates, which are a group of secondary metabolites that are only present in Brassicaceae and immediate families. They have various functions within the plant, being especially important in the defense against pathogens and herbivores (Poveda et al. 2020).

Many *Brassica* crops are of great economic significance, as they are cultivated as vegetables, oilseed sources, condiments, and forages. Climate change, pathogen variation, and inappropriate farming methods, such as continuous and high-intensity cropping, contribute to disease outbreaks, which pose threats to crucifers "production." Various biotic and abiotic stresses cause production losses (Saharan et al. 2017). Traditional approaches for disease prevention include agricultural, physical, chemical, and biological controls, and integrated pest management (IPM) strategies. Physical approaches, such as high-temperature treatment and light trapping, chemicals, such as fungicides and bactericides, and biological agents, such as *Bacillus subtilis* and arbuscular mycorrhiza, are frequently used. IPM has been extensively studied and can achieve some effect for certain diseases. However, the approaches are often complicated, costly, and/or environmentally damaging. In contrast, natural resistance in crucifer's hosts is the most desirable strategy and could be integrated with other approaches for high-efficiency disease control. Two types of plant immunity have been identified to date: pathogen/microbe-associated molecular pattern (PAMP/MAMP)-triggered immunity, which is activated by cell surface-localized pattern recognition receptors by the recognition of PAMPs/MAMPs, and effector-triggered immunity activated by host resistance (R) genes through the recognition of pathogen-specific effector molecules, which is in accord with the gene-for-gene theory. Most R genes identified to date encode nucleotide-binding leucine-rich repeats (NB-LRRs), including coiled-coil NB-LRRs (CC-NB-LRRs) and Toll interleukin 1 receptor NB-LRRs (TIR-NB-LRRs). Moreover, some

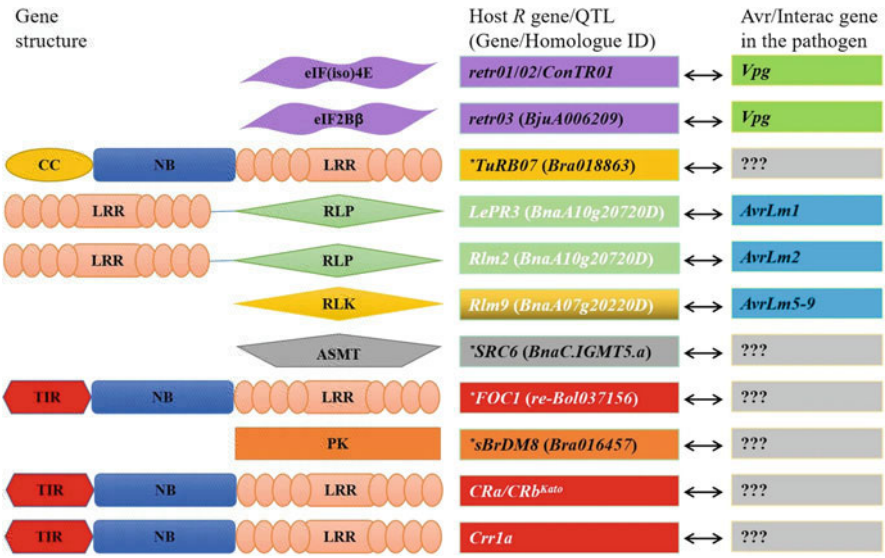


Fig. 1.1 Resistance genes identified in *Brassica* crops and their avirulence/interactor genes in the pathogens. ASMT, *N*-acetylserotonin Omethyltransferase; CC, coiled-coil domain; eIF (iso) 4E, eukaryotic translation initiation factor isoform 4E; eIF2Bβ, eukaryotic translation initiation factor 2Bβ; LRR, leucine-rich repeat; NB, nucleotide-binding domain; PK, protein kinase; RLP, receptor-like protein. *Putative genes that have not been functionally validated. ??? The avirulence or interaction genes in the pathogens that have not yet been characterized (Lv et al. 2020)

R genes encode receptor-like kinases (RLKs), transmembrane receptor-like proteins (RLPs), cytoplasmic kinases, and proteins with a-typical molecular motifs. Various R genes with flexible molecular mechanisms provide powerful weapons that protect the plant host from pathogens.

Many R genes have been identified and successfully applied to improve *Brassica* crop resistance against various diseases, which not only ensures *Brassica* production but also facilitates the discovery of host–pathogen interactions. Moreover, the genomic era characterized by massive genome and omic data has made fast and accurate R gene studies possible. The release of the reference genome data of the six *Brassica* species in addition to *B. carinata* has provided vital information for determining the genetic and molecular basis of disease resistance. Since the 2010s, researchers have performed extensive, high-quality genomic, postgenomic, and omics studies in *Brassica* species and have discovered a variety of R genes and closely related genes, which not only provide further insight into the resistance molecular mechanism and host–pathogen co-evolutionary arms race but also facilitate accurate molecular breeding at the whole-genome level (Lv et al. 2020; Fig. 1.1).

1.3 The State of *Brassica* Genomics

The genomes of diploid *Brassica* species are three to five times the size of the *Arabidopsis* genome, ranging from 0.97 pg/2C (468 Mb/1C; where C is haploid DNA per nucleus), for *B. nigra* to 1.37 pg/2C (662 Mb/1C) for *B. oleracea* 696 Mb. The amphidiploid genomes, which have two sets of chromosomes from each parent species, range from 2.29 pg/2C (1105 Mb/1C) to 2.56 pg/2C (1235 Mb/1C). Relationships among the three diploid species—*B. rapa* (syn. rapa; $2n = 20$, genome AA), *B. nigra* ($2n = 16$, genome BB), and *B. oleracea* ($2n = 18$, genome CC)—and three amphidiploids *B. napus* ($2n = 38$, genome AACC), *B. juncea* ($2n = 36$, genome AABB), and *B. carinata* ($2n = 34$, genome BBCC) are well known. Rapid-cycling strains of *B. oleracea* have been developed that have a life cycle as short as that of *Arabidopsis thaliana*, making them easier to study (Nagaharu 1935; Williams and Hill 1986; Paterson et al. 2001).

Neutral DNA polymorphisms are common among coeno-specific genotypes of most *Brassica* species, and at least 15 molecular maps have been produced, using at least 900 different publicly available *Brassica* and *Arabidopsis* DNA probes (which are each used with similar efficacy). Virtually all the probes hybridize to two or more loci, even in diploid *Brassica* species; however, only a subset of loci segregates for allelic variation in any one mapping population. The subset of genetically mapped loci can, therefore, be used to infer the locations of many additional sequence tagged sites (STSs) (Lan et al. 2000). The comparative organization of the chromosomes of *Brassica* and *Arabidopsis* is well studied. On the basis of 186 corresponding loci, about 19 chromosome structural rearrangements differentiate *B. oleracea* and *A. thaliana* orthologs, suggesting that chromosomal tracts of about 20–25 cM remain largely co-linear in the two taxa (Lan et al. 2000). Microsynteny studies involving comparative genetic and physical mapping of specific chromosome segments have shown largely conserved gene order in *Arabidopsis* and *Brassica*, but some disruption in gene content by deletions or insertions. Furthermore, duplicated partial gene clusters are commonly found in both *Arabidopsis* and *Brassica* species. Comparative sequencing has permitted orthology assignment of some duplicated segments in *Arabidopsis* and *Brassica* species (Wroblewski et al. 2000).

1.4 Crucifers' Host Genome

The members of the crucifer family are too diverse and the genomic structures are very complicated. In general, the cultivated *Brassica* species as per the triangle of U including diploid as well as allotetraploids species, the diploid possess the most variability while there is less variability, as compared to the diploid, in the amphidiploid species, as it is derived from the diploid species with the few chance crosses followed by the amphidiplodization. Also the selection pressure on the amphidiploid species is more due to its nucleo-cytoplasm and environmental instability. But still a considerable genetic diversity is present within these three amphidiploid species (Song et al. 1998). Based upon studies of genetic diversity, *B. napus* may be

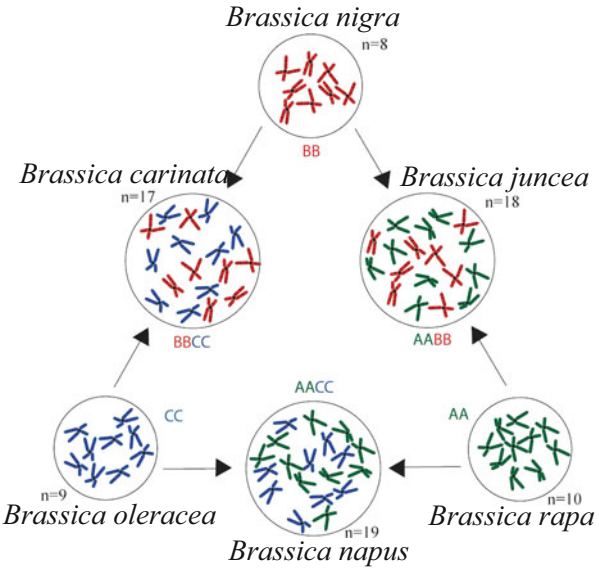


Fig. 1.2 Representing inter-relationship among the diploid and amphidiploid species in triangle of U by Nagaharu (1935)

considered as the most ancient amphidiploid, succeeded by *B. juncea* and *B. carinata*. Two major factors are responsible for general diversity within amphidiploids, multiple hybridizations with different diploid parents and genome modifications following polyploidization. Early comparative studies conducted at the level of genetic linkage maps revealed extensive duplication within *Brassica* genomes (Lagercrantz and Lydiate 1996) and tracts of co-linearity disrupted by multiple rearrangements between the genomes of *B. nigra* and *A. thaliana* (Lagercrantz 1998). The genome of different U-triangle species along with *Arabidopsis thaliana* (Figs. 1.2 and 1.3) is as follows.

1.4.1 *Arabidopsis thaliana*

The flowering plant *Arabidopsis thaliana* is an important model system for identifying genes and determining their functions. So, *A. thaliana* made a large contribution to our molecular understanding of key concepts in biology by successful integration of different fields of molecular biology research. The chromosome number of *Arabidopsis thaliana* is $2n = 10$ and it has a polyploid origin, although there has been a dramatic genome reconstruction and loss of genes from the duplicated genome segments (Arabidopsis Genome Initiative 2000; Blanc et al. 2000; Ku et al. 2000; Mayer et al. 2001). The *Brassica* and *Arabidopsis* lineages diverged 20 MYA (Yang et al. 1999). Phylogenetic analysis groups of the *Brassica* species into the *nigra* and *rapa/oleracea* lineages (Warwick and Black 1991), which diverged 8 MYA (Lysak et al. 2005). With the progression of the *Arabidopsis* genome project during the late 1990s, thousands of "interesting" genes were

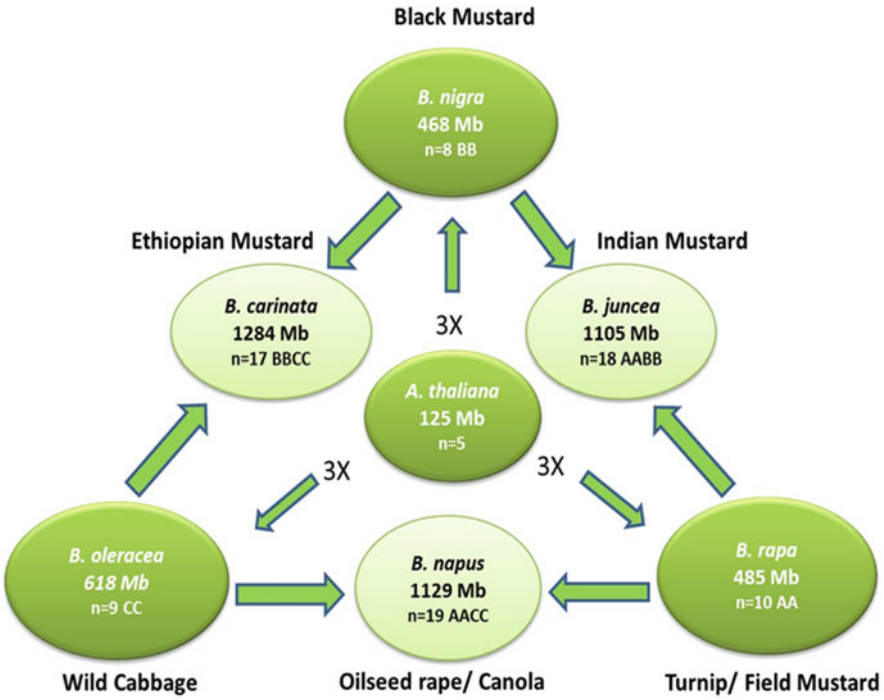


Fig. 1.3 U's triangle of *Brassica* species with their estimated genome size. The inferred relationship with the model species *A. thaliana* is indicated, but it should be noted that each of the species shown share a common ancestor with an increased chromosomes number

discovered at an unprecedented pace, fueling the desire by plant researchers to test their favorite hypothesis about these genes' potential functions. The genome size of *Arabidopsis thaliana* is 125 Mbp and genome contains 25,498 genes encoding proteins from 11,000 families (Arabidopsis Genome Initiative 2000).

1.4.2 *Brassica carinata*

Brassica carinata (BBCC, $n = 17$) was formed from *B. oleracea* (CC, $n = 9$) and *B. nigra* (BB, $n = 8$). This species is characterized by the slow steady growth of *B. oleracea* and the mustard oil content of *B. nigra*. Wild forms of *B. carinata* are not known, but primitive domesticated types are cultivated in upland areas of Ethiopia and further south into Kenya. This hybrid may have originated from kale land races of *B. oleracea* types hybridizing with wild or semi-domesticated forms of *B. nigra*. Both kale and *carinata* crops thrive in cool environments typical of the Ethiopian plateau. The local farmers grow them in "kale gardens." This term, translated into many languages and dialects, is commonly found throughout those rural areas using vegetables derived from *B. oleracea* types. *Carinata* crops themselves are

alternatively named "guomin," Abyssinian mustard or Ethiopian cabbage, and provide leafy vegetables and sources of oils. Cabbage itself can be a ubiquitous term used to describe cole *Brassicas* and not necessarily synonymous with the sophisticated heads seen on today's super market shelves. The flow cytometry estimation of *B. carinata* gave a picture of genome size around 642 Mb (Johnston et al. 2005).

1.4.3 *Brassica juncea*

Brassica juncea ($n = 18$) is a hybrid between *B. rapa* ($n = 10$) and *B. nigra* ($n = 8$) (Frandsen 1943) producing large leaves and with the rapid growth of *B. rapa* and the mustard oil of *B. nigra*. Use of *B. juncea* as a source of vegetable oil is gaining importance in India; while throughout Asia, especially in China and Japan, the plant has a great diversity of cultivated forms used as staple vegetables of immense dietary importance. Reputably, wild forms are still found on the Asia Minor plateau and in southern Iran. The genome size of *B. juncea* variety Tumida is around 922 Mb (Yang et al. 2016).

1.4.4 *Brassica napus*

Brassica napus ($n = 19$), has wild forms in Sweden, Denmark, The Netherlands, and the UK. It developed from hybridization between *B. rapa* ($n = 10$) and *B. oleracea* ($n = 9$) followed by chromosome doubling. This hybrid may have formed as *B. oleracea* types expanded their range along the coasts of northern Europe and *B. rapa* extended from the Irano-Turanian regions. Alternatively, *B. napus* may have Mediterranean origins or, as seems likely, there were several centers of evolution. The wild populations of *B. napus* have acquired major scientific significance recently as they present means of determining the potential for gene flow to and from genetically modified cultivars of oilseed rape. The sequence analysis reveals the genome size of winter *B. napus* is 925 Mb (Lee et al. 2020).

1.4.5 *Brassica nigra*

Brassica nigra (black mustard), itself the ancestor of culinary mustards, is found widely and distributed as annual herbs growing in shallow soils around major rocky Mediterranean coasts. The *B. nigra* is the diploid species and also progenitor of the two allotetraploids species viz., *B. juncea* and *B. carinata*. The chromosome numbers of *B. nigra* is 16 with BB genome. The estimated genome size as per the flow cytometry analysis is 632 Mb (Johnston et al. 2005). The sequence of *B. nigra* variety Sangam using Illunima HiSeq platform depict the genome size of *B. nigra* as 591 Mb (Yang et al. 2016).