Devendra K. Choudhary · Manoj Kumar
Ram Prasad · Vivek Kumar *Editors*

In Silico Approach for Sustainable Agriculture

Springer

In Silico Approach for Sustainable Agriculture

Devendra K. Choudhary • Manoj Kumar
Ram Prasad • Vivek Kumar
Editors

In Silico Approach for Sustainable Agriculture

Springer

Editors
Devendra K. Choudhary
Department of Microbiology, Faculty of Life Sciences
PDM University
Bahadurgarh, Haryana, India

Ram Prasad
Amity Institute of Microbial Technology
Amity University
Noida, Uttar Pradesh, India

Manoj Kumar
Centre for Life Sciences
Central University of Jharkhand
Ranchi, Jharkhand, India

Vivek Kumar
Amity Institute of Microbial Technology
Amity University
Noida, Uttar Pradesh, India

ISBN 978-981-13-0346-3 ISBN 978-981-13-0347-0 (eBook)
https://doi.org/10.1007/978-981-13-0347-0

Library of Congress Control Number: 2018944672

Printed on acid-free paper

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

Preface

From recently *in silico* approach in biological sciences and agriculture is an interdisciplinary science developed by deploying benign use of computer, statistics, biology, and mathematics to analyze genome arrangement and contents and biological sequence data, and predict the structure and function of macromolecules that use in interpreting and decoding plant genome. The broad amount of data produced in life sciences resulted in the evolution and development of bioinformatics. Omics, bioinformatics, and computational tools are very essential to understand genomics and the molecular systems that underlie several plant functions. Various new omic layers such as genome, hormonome, metabalome, interactome, and epigenome analysis have emerged by technological advances. Such integration of information enables and facilitates the identification of expression of gene which helps to interpret the relationship between phenotype and genotype, thus approving from genome to phenome system-wide analysis. Earlier biological research that used laboratories, plant clinics and field is now at *in silico* or computer level (computational). Bioinformatics develops software, algorithms, databases, and tools of data analysis to make discoveries and infer the information. Application of various bioinformatics tools and databases enables analysis, storage, annotation, visualization, and retrieval of outcomes to help enhance understanding in living system research. Thus, it will help to improve the plant quality based on health care disease diagnosis. In this book we describe the bioinformatics approaches (databases and tools) in plant science and implication of next-generation sequencing (NGS) technology on crop genetics. The proposed book will be benign to researchers involved in sustenance of agriculture with below described points:

- It has inclusion of *in silico* characterization of microbes deployed for soil fertility.
- Chapters describe *in silico* deployment of benign and pathogenic microbial strains relatedness.
- Glimpses given for *in silico* characteristics of microbial and plant genes.
- Emphasis given on biotechnological perspectives of *in silico* deployment in agriculture.

- Overall, this book describes role of *in silico* approaches deployed for sustenance of agriculture.

For inclusion of *in silico* deployment for microbial and potent gene characterization, initially, our research group has screened potent bacterial isolates capable of having plant growth promoting activity and can elicit induced systemic resistance in plant. Further molecular characterization of 16 s rRNA followed by plant growth promoting (PGP) genes was done to confirm the identity and activity.

For Gram +ve Bacterium

PCR Amplification of 16S Ribosomal RNA Gene

PCR amplification of the 16S rRNA gene from genomic DNA of the bacterium isolate SJ-5 was done by using universal 16S rRNA primers: *27F*-AGAGTTTGATCM TGGCTCAG and *1492R*-CGGTTACCTTGTTACGACTT in thermal cycler (Kyratec, Australia). PCR program and master mix ingredients concentration used for the amplification are given in the Table 1. For the assessment of gene, PCR product was loaded in the 1% agarose gel (wt/vol) containing EtBr and run for 45 min at 85 Volts in 1X TAE buffer. Observation for the DNA band was done by visualizing gel for the DNA band in UVITECH gel doc system.

PCR Product Purification and Sequencing

Obtained PCR production was purified using Promega kit "Wizard® SV Gel and PCR Clean-Up System" as per manufacturer's instruction. Purified PCR product was checked quantitatively and qualitatively through biophotometer and gel electrophoresis, respectively, and Sanger sequencing was done by Shrimpex Biotech Services Pvt. Ltd., Chennai, using BDT v3.1 Cycle Sequencing Kit on ABI 3500 Genetic Analyzer.

Table 1 PCR ingredients for 16S rRNA gene amplification

PCR master mix ingredients	Volume used for 20 µl reaction mixture
10X *Taq* polymerase buffer	2 µl
25 mM $MgCl_2$	2 µl
10 mM dNTP mix	2 µl
Forward primer (20 µmole)	1 µl
Reverse primer (20 µmole)	1 µl
Taq DNA polymerase (3 U/µl)	0.4
DNA template (100 ng/µl)	1 µl
NFW	10.6 µl

16S rRNA Gene Sequence Homology and Phylogeny Analysis

Molecular identification of the bacterium SJ-5 was done by homology analysis of 16S rRNA gene sequence obtained through sequencing, using BLAST tool of NCBI available at http://blast.ncbi.nlm.nih.gov/. Based on maximum identity score first ten sequences were selected for further analysis. Selected sequences were aligned using Clustal W and checked for the gap. Phylogenetic and molecular evolutionary analysis was performed using MEGA 6 software (Tamura et al. 2013).

Sequence Submission and Culture Deposition

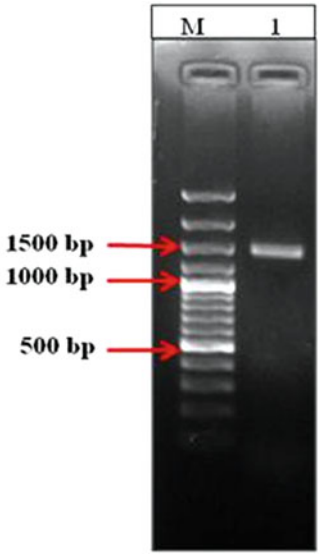
The nucleotide sequence of the 16S rRNA gene was submitted to NCBI GenBank, and accession number provided by the NCBI is KJ 184312. Bacterial culture was also deposited to Microbial Culture Collection (MCC), Pune, with accession number "MCC-2069" (Table 2)

Molecular Characterization of Plant Growth Promoting Bacterium and Phylogeny Analysis

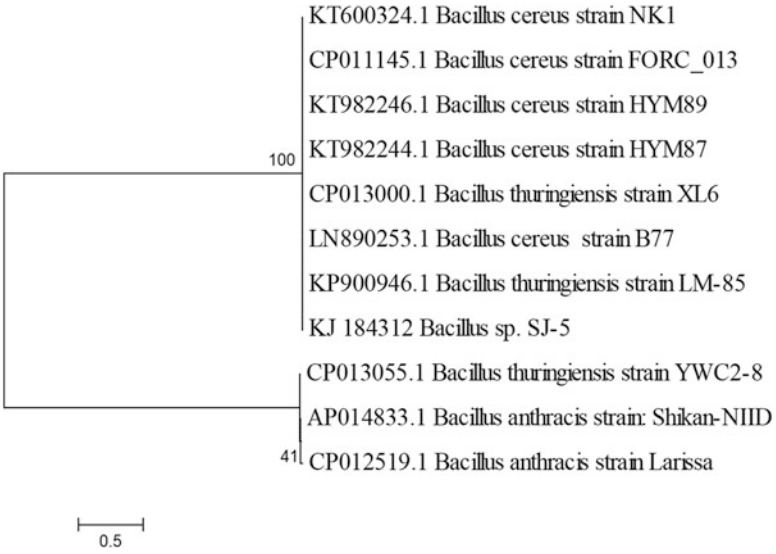
Molecular characterization of plant growth–promoting bacterium SJ-5 was done by amplifying 16S rRNA gene from genomic DNA by using universal 16S rRNA primers. On the agarose gel, a sharp band around 1.5 kb was observed. Sequence homology and phylogenetic analysis of the obtained sequence show similarity of the strain with *B. cereus* and *B. thuringiensis,* and hence, the bacterium was confirmed as a member of genus *Bacillus* and submitted by the name of *Bacillus* sp. SJ-5 in the NCBI GenBank. Accession number provided by NCBI to bacterium is KJ 184312. Bacterial culture was also deposited to Microbial Culture Collection (MCC), Pune, with accession number MCC-2069.

Table 2 PCR programming for 16S rRNA gene amplification

PCR steps	Temperature (°C)	Duration (min)
Initial denaturation	94	4
Starting of loop × 30 cycle		
Denaturation	94	1
Primer annealing	55	1
Extension	72	1.30
Closing of loop		
Final extension	72	10



PCR amplification of 16S rRNA gene of *Bacillus* sp. SJ-5; Lanes M-Marker (100 bp plus, Thermo scientific); 1-16S rDNA gene.

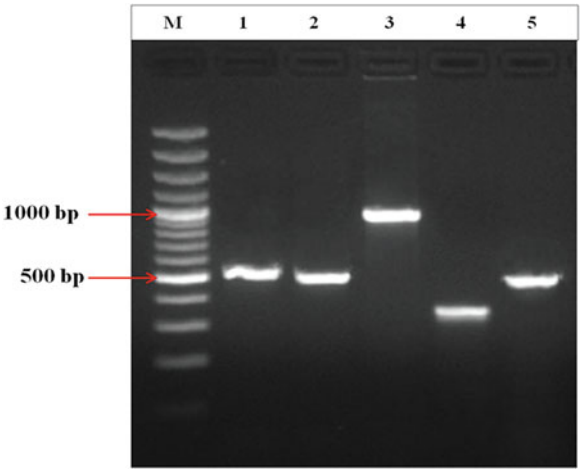


Evolutionary Relationships of Bacillus *sp. SJ-5*

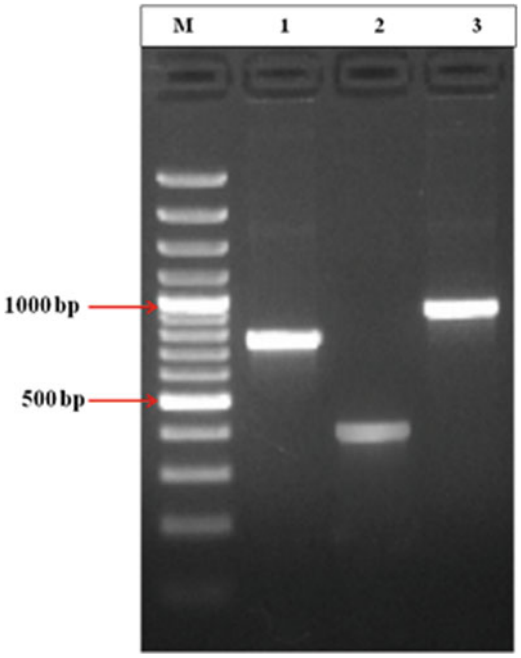
The evolutionary history was inferred using the neighbor-joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The evolutionary distances were computed using the maximum composite likelihood method.

Molecular Characterization of Plant Growth Promoting Properties of Bacteria

Molecular characterization of plant growth–promoting activities of bacterium SJ-5 was done by amplifying plant growth–promoting and biocontrol genes of SJ-5 by using GSP. Gene amplification using GSP showed sharp bands of the specific genes near to desired amplicon size. Sequence homology analysis of the concerned gene confirmed the amplification of the desired gene.



PCR amplification of plant–growth promoting genes of *Bacillus* sp. SJ-5; Lanes M-Marker (100 bp plus, Thermo scientific); 1-*Glucose-1-dehydrogenase*; 2-*ACC deaminase*; 3-*Tryptophan 2-monooxygenase*; 4-*Ferredoxin-nitrite reductase*; 5-*Siderophore biosynthesis gene.*



PCR amplification of biocontrol genes of *Bacillus* sp. SJ-5; Lanes M-Marker (100 bp plus, Thermo scientific); 1- *β-1,3-glucanase*; 2- *Chitinase*; 3-*Zwittermicin* (Figs. 1 and 2)

For Grame –ve Bacterium

Gene Amplification

Seven PGP properties genes were used for functional characterization, namely ACC-deaminase (*acds*), tryptophan-2-monooxygenase for IAA production (*IaaM*), glucose-6 phosphate dehydrogenase for gluconic acid production (*g6pd*), siderophore (*sid*), alpha amylase (*amy*), nitrite reductase (*nr*), and proline (*p5cr*). Genes were amplified by using their specific primers that were designed by IDT oligo-analyzer software and synthesized from Sigma-Aldrich (India). The 16S rRNA gene was amplified using universal primers. The PCR reaction mixture for 25 µL was given in Table 3.

The PCR amplification was carried out in a gradient thermal cycler (Bio-Rad). The condition of PCR cycle was given in Table 4.

Amplified PCR amplicons were resolved by electrophoresis in 1x TAE running buffer using a 1.2% agarose gel containing ethidium bromide (0.5 µg/mL) (Sambroock et al. 1989) and visualized on a gel documentation system (Bio-Rad Laboratories, CA, USA) to confirm the expected size compared with DNA marker.

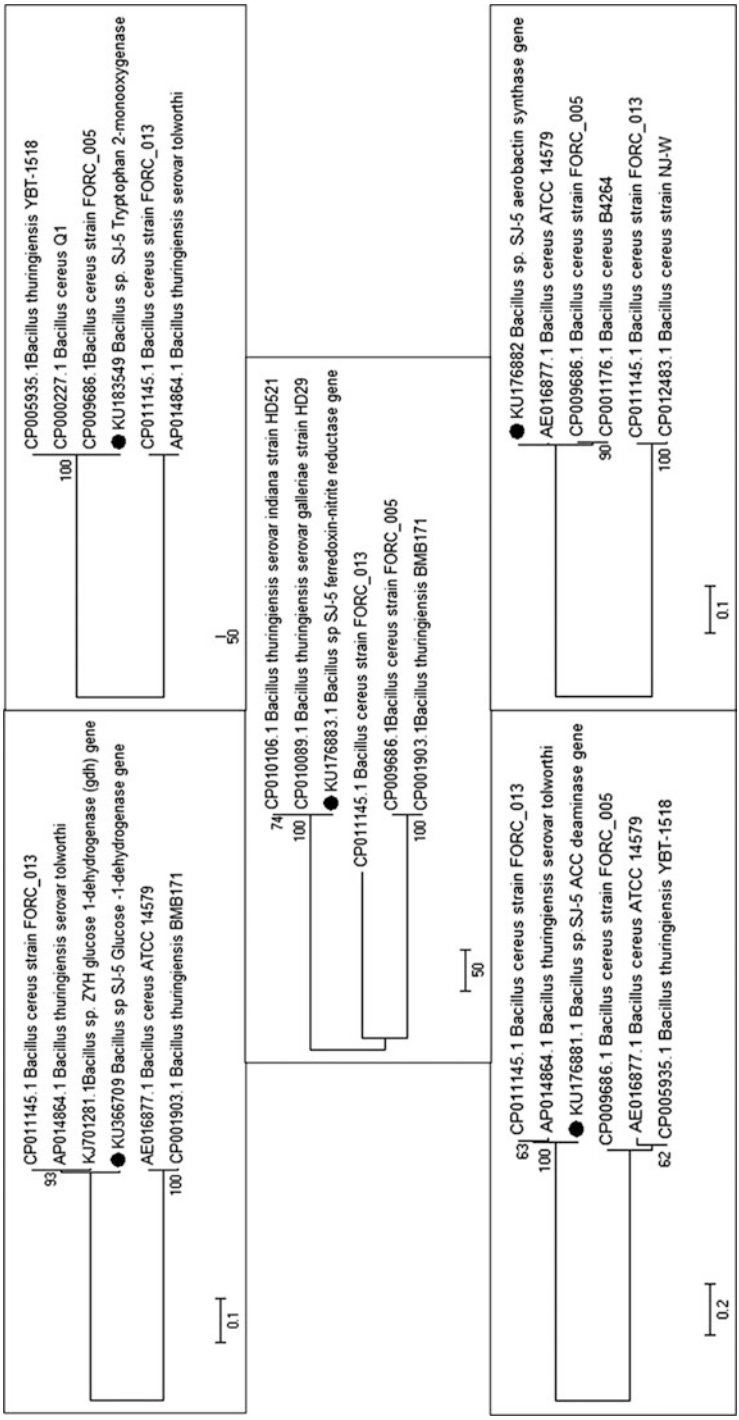


Fig. 1 Phylogenetic analysis of plant growth–promoting genes of *Bacillus* sp. SJ-5

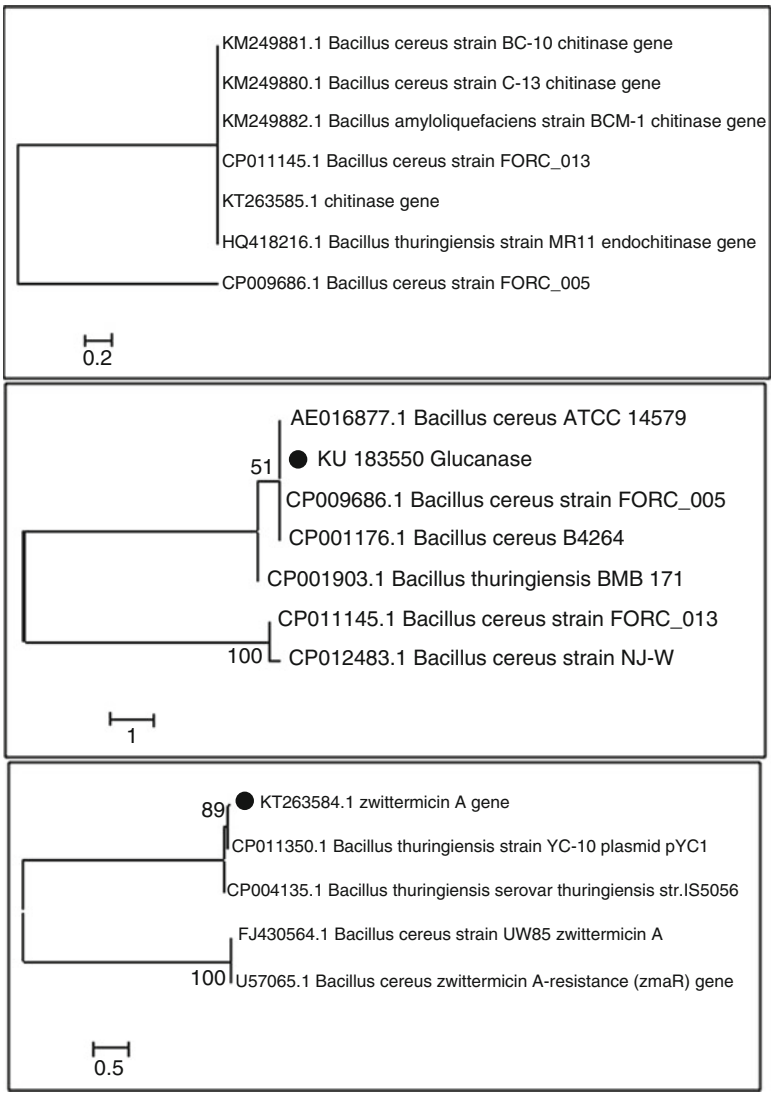


Fig. 2 Phylogenetic analysis of biocontrol genes of *Bacillus* sp. SJ-5

Table 3 PCR ingredients for 25 µL reaction mixture

Ingredients	Volume in 25 µL PCR mixture
10X *Taq* polymerase buffer	2.5
2.5 mM dNTP mix	1.0
25 mM $MgCl_2$	1.5
20 pmol forward specific primer	0.7
20 pmol reverse specific primer	0.7
Taq DNA polymerase 5 U	0.2
DNA template 100 ng	1.0
PCR grade water	17.4

Table 4 PCR condition for gene amplification

PCR steps	Temperature	Time duration
Initial denaturation	95 °C	5 min
35 cycles		
Denaturation	95 °C	30 s
Primer annealing	***	30 s
Elongation	72 °C	40 s
Final extension	72 °C	7 min

***The details regarding primer annealing temperature of each gene are given in next Table 5

Table 5 Primer details of genes tested in the present study

Gene	Primer sequences (forward and reverse)	Annealing temperature (°C)
acds	F- 5′-AACCACCAAGCGTCGTAATC-3’	58
	R- 5′-GGCAACAAGTGGTTCAAACTC-3′	
p5cr	F- 5′-CTTTATCTCTGAGCACCTCCAG-3′	62
	R- 5′-CATGAACGGCTGGTTCTATTTC-3′	
IaaM	F- 5′-GACTTCCCCAACTCGATGCTG-3′	62
	R- 5′-ATCCACATCTTTTGCGAGAACAG-3′	
g6pd	F- 5′-ACAAACAGGTTCTGATTGCCG-3′	58
	R- 5′-TGGGGCTATTTCGACAAGGC-3′	
amy	F- 5′-ACTTCTGGCACCGTTTCTAC-3′	58
	R- 5′-GCGTAGTAGTTCCACAGGTAATC-3′	
nr	F- 5′-TGGTGACGTTATGGCAAGAG-3′	58
	R- 5′-CACTACCGTTACCGCATGAA-3′	
sid	F- 5′-CCATTGCATTAGGTCCAGAAATG-3′	60
	R- 5′-GCCAATGCCAATGTGGATTAC-3′	
16 s-rDNA	F- 5′-AGAGTTTGATCCTGGCTCAG-3′	60
	F- 5′-AAGGAGGTGATCCAGCCGCA-3′	

Gene Sequencing and Analysis

The amplified PCR products were purified and both strands were sequenced using respective forward and reverse primers. The nucleotide sequences were di-deoxy cycle sequenced with fluorescent terminators (Big Dye, Applied Biosystems) and run in ABI 3730xl DNA Analyzer (Applied Biosystem, USA). Sequence analyses were performed with Chromas-Pro software in order to verify sequence quality and then compared with nucleotides database provided by the National Center for Biotechnology Information using the BLAST (Basic Local Alignment Search Tool). The alignment scores and the percent sequence identity were determined for the closest identity of the sequences obtained. Identification to the species level was determined as a 16S rDNA sequence similarity with a prototype strain sequence in the GenBank.

Phylogeny Based on 16 rRNA Gene Analyses

About 1.5 kb fragment of 16S rRNA gene of the AU isolates was sequenced. The sequences obtained were analyzed to determine the phylogenetic position of the AU strain. Sequence database searches (GenBank and EzTaxon) revealed that the AU strain was phylogenetically most closely related to members of the genus *Pseudomonas* (Fig. 3). Phylogenetic tree obtained using the neighbor-joining methods

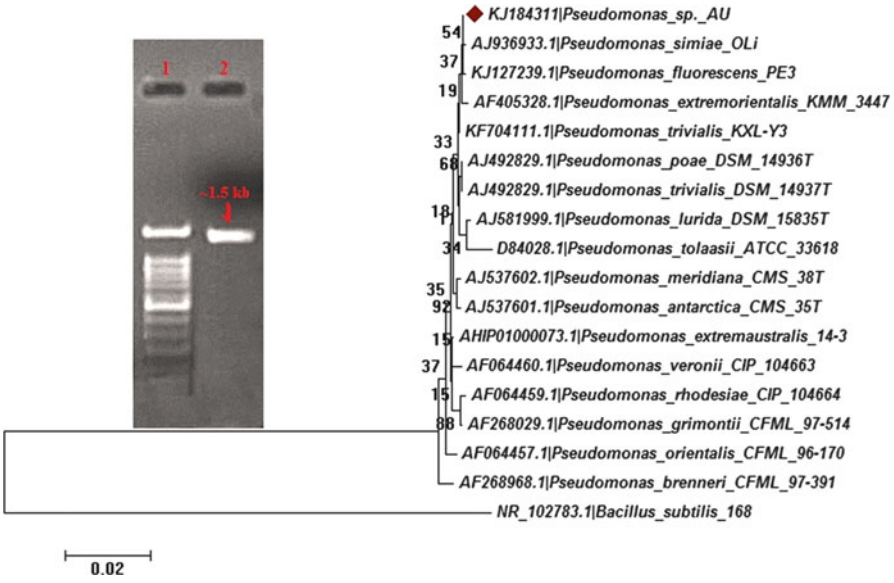


Fig. 3 Construction of phylogenetic tree based on 16S rRNA gene sequencing by neighbor-joining method

revealed that AU strain showed a similarity of 99.93% with *Pseudomonas simiae* OLi type strain (AJ936933) with "*P. fluorescens* intrageneric cluster" and placed the bacterial strain on a separate branch within this intrageneric cluster containing *P. fluorescens* PE3 type strain (99.85%; KJ127239). The other species most closely related to AU strain was *Pseudomonas extremorientalis* KMM3447 (99.63%; AF405328). *Bacillus subtilis* 168 (NR102783) was taken as an out group for the analysis. The AU strain 16S rRNA gene sequence (1380 bp) was submitted to NCBI with KJ184311 accession number, and the culture was deposited in Microbial Type Culture Collection (MTCC), IMTECH, Chandigarh, with accession number 12057.

The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. The analysis involved 18 nucleotide sequences in which one is from outer group *Bacillus*. Values shown next to the branches are the percentage of replicate trees with associated taxa clustered together in the bootstrap test (1000 replicates). Evolutionary analyses were conducted in MEGA 6.

Phylogeny Analysis of PGP Properties Genes

Gene amplification using specific gene primers showed sharp bands for *acds* (410 bp), *IaaM* (330 bp), *g6pd* (630 bp), sid (360 bp), tre (510 bp), nr (630 bp), and p5cs (250 bp) (Fig. 4). All genes were sequenced and submitted in NCBI; GenBank accession numbers are listed in Table 6. As expected, all PGP genes of AU bacterial strain tested showed similarity to the *P. fluorescence* intrageneric cluster. Regarding the phylogeny based on the *IaaM* and *nr* gene sequences, the AU

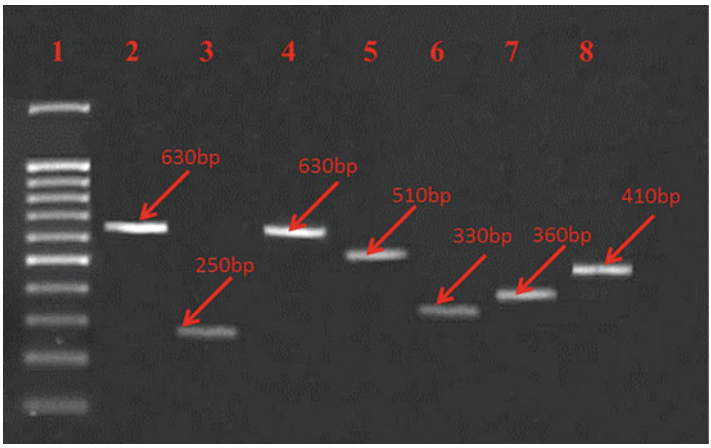


Fig. 4 PCR amplification of seven different PGP genes in AU bacterial isolate
Lane 1- DNA marker (100 bp), Lane 2- *g6pd*, Lane 3- *p5cr*, Lane 4- *nr,* Lane 5- *amy*, Lane 6- *IaaM*, Lane 7- *sid,* Lane 8- *acds*

Table 6 Details of gene sequences submitted in NCBI

NCBI accession number	Details	Bacterial strain
KU159726	ACC-deaminase (*acds*)	*Pseudomonas simiae* AU
KT805422	Nitrite reductase (*nr*)	*Pseudomonas simiae* AU
KT805423	Alpha amylase (*amy*)	*Pseudomonas simiae* AU
KT805424	Indole acetic acid (*IaaM*)	*Pseudomonas simiae* AU
KU204777	Siderophore (*sid*)	*Pseudomonas simiae* AU
KU204778	Gluconic acid (*g6pd*)	*Pseudomonas simiae* AU
KU204779	Proline (*p5cr*)	*Pseudomonas simiae* AU
KJ184311	16S ribosomal RNA gene	*Pseudomonas* sp. AU

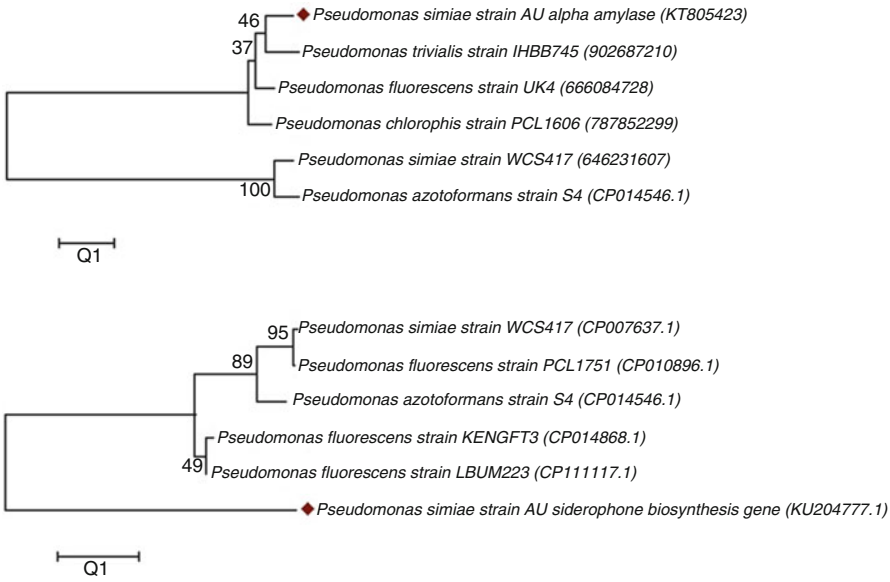


Fig. 5 Phylogenetic analysis of *acds* and *sid* gene of AU bacterial isolate

bacterial strain forms an independent cluster, which includes *P. simiae* WCS417 (CP007637), whereas in *acds*, *g6pd*, *sid*, *tre*, and *p5cs* genes phylogeny, AU bacterial strain was included with *P. fluorescens* PICF7 (CP005975) in an independent cluster. All other species from the genus *Pseudomonas* are found outside this cluster (Fig. 5).

The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. Values shown next to the branches are the percentage of replicate trees with associated taxa clustered together in the bootstrap test (1000 replicates). Evolutionary analyses were conducted in MEGA 6 (Fig. 6).

The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. Values shown next to

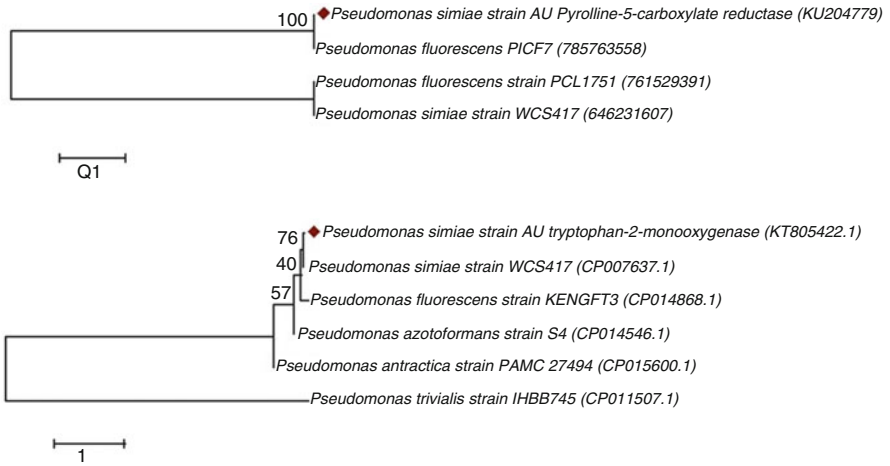


Fig. 6 Phylogenetic analysis of *p5cr* and *IaaM* isolated from AU bacterial isolate

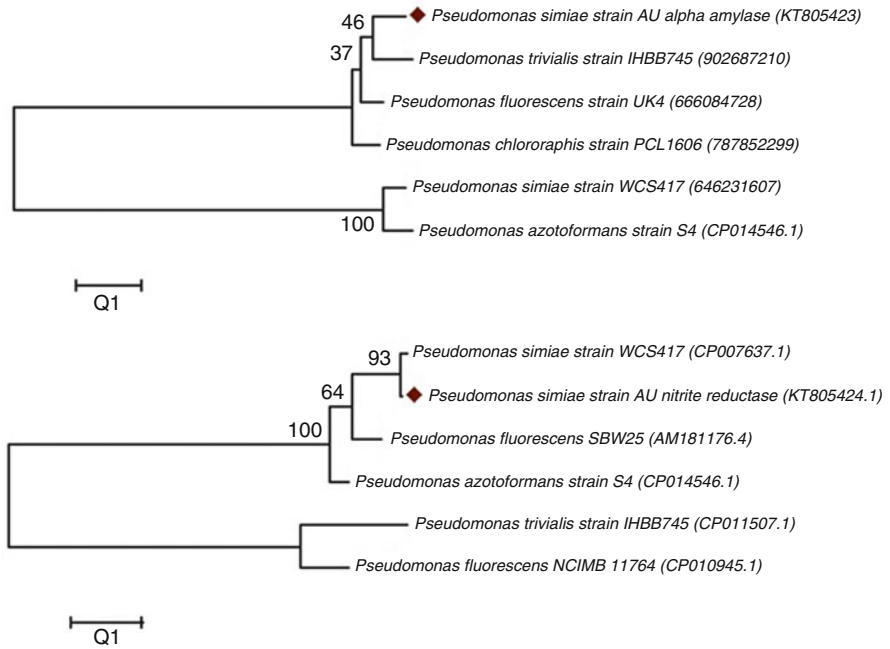


Fig. 7 Phylogenetic analysis of *amy* and *nr* isolated from AU bacterial isolate

the branches are the percentage of replicate trees with associated taxa clustered together in the bootstrap test (1000 replicates). Evolutionary analyses were conducted in MEGA 6 (Fig. 7).

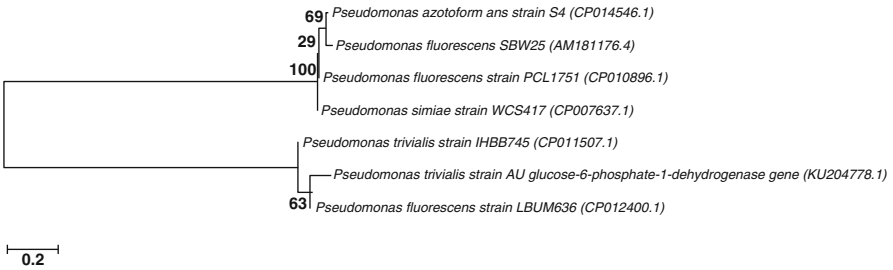


Fig. 8 Phylogenetic analysis of *g6pd* isolated from AU bacterial isolate

The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. Values shown next to the branches are the percentage of replicate trees with associated taxa clustered together in the bootstrap test (1000 replicates). Evolutionary analyses were conducted in MEGA 6 (Fig. 8).

The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. Values shown next to the branches are the percentage of replicate trees with associated taxa clustered together in the bootstrap test (1000 replicates). Evolutionary analyses were conducted in MEGA 6.

All isolates were screened based on physiological and plant growth promotion properties. One bacterial isolate AU was found to produce IAA, ACC-D, siderophore, exopolysaccharide, Pi solubilization, and beneficial VOCs production under 10% NaCl condition. Further, AU isolate was selected for induced systemic tolerance study and subjected to biochemical and molecular characterization. Based on the morphological and biochemical properties, AU isolate was tentatively identified as *Pseudomonas* sp. After molecular characterization with 16S rRNA gene, AU strain showed a similarity of 99.93% with *Pseudomonas simiae* OLi type strain (AJ936933) with "*P. fluorescens* intrageneric cluster." In addition, PCR amplification was confirmed in the presence of ACC-deaminase, IAA production, gluconic acid, siderophore, alpha amylase, nitrite reductase, and proline gene in bacterial isolate AU. All gene sequences of AU isolate were submitted to NCBI accession number, and the culture was deposited in an IDA approved culture collection, Microbial Type Culture Collection (MTCC), IMTECH, Chandigarh.

In the present book, editors compiled researches carried out by researchers in the form of compendium with elaborate description related to "*in silico* deployment for characterization of microbial and plant genes for sustenance of agriculture."

Chapter 1 emphasizes overview of the retrobiosynthesis methodology and summarize about the available genome mining tools utilized in the forward and retrosynthetic approaches to envisage secondary metabolite biosynthetic gene clusters and vice versa as well as pathway design analysis and evolution.

Chapter 2 describes the *in silico* role of PGPR that are associated with the rhizosphere of soybean grown in semiarid areas of Rajasthan. We also sought to identify and characterize representative PGPR with respect to growth-promoting attributes and studied their salinity tolerance.

Chapter 3 highlights the homogeneity in soils for methanogenic diversity, although methane production potential varied, possibly due to the presence of methanogens in different proportions and difference in the soil characteristics.

Chapter 4 has given emphasis on *in silico* spread and transmission of Geminiviruses back to crop plants which enhances the host range of these viruses. Thus, there is a pressing need for additional information on the diversity and distribution of Geminiviruses in ornamental plants.

Chapter 5 elaborately describes a survey of the available *in silico* approaches to identify the candidate genes conferring disease resistance in plants. After providing a brief overview of the multilayered defense mechanism, the chapter discusses different approaches for the step wise identification of disease-resistant candidate genes in plants.

Chapter 6 focuses on molecular tools to gain insight into the mechanism of overcoming heat stress, and desiccation by expression of heat shock proteins and blue light sensitivity on nodulation that has added new dimensions to this area of research.

Chapter 7 describes *in silico* approach to how drought affects the soybean production worldwide and utilize specific bradyrhizobial strains to confer tolerance to soybean plants under drought stress and to understand the mechanisms imparting in the reduction of abiotic stress e.g., GOGGAT MAPK, different polysaccharide and other precursors involved in drought stress recovering mechanism.

Chapter 8 provides the comprehensive overview on the *in silico* approach for the reconstruction of biochemical pathway and different databases and computational tools associated with it.

Chapter 9 deals with the association of modern techniques like nano technology which can be linked further with agricultural problems to elevate conditions like sustainability and competitiveness of organic markets.

Chapter 10 describes the various *in silico* tools that comprised of databases and software and assist to reduce the "sequence-function gap" and help in the broad spectrum study of soil microorganisms and their application toward sustainable agriculture.

Chapter 11 describes how *in silico* tools have helped to explore the microbiota of the soil and how these can be used as a guiding line to innovate new agricultural norms for sustainable environment.

Chapter 12 emphasizes *in silico* secondary metabolites produced by plants and their potential applications along with the databases which have been made to easily retrieve the required data about them for scientific and academic purposes.

Chapter 13 highlights on *in silico* molecular docking studies to investigate the binding interactions between natural compounds and 13 various anti-Alzheimer drug targets. Three known cholinesterase inhibitors (Donepezil, Galantamine,

and Rivastigmine) were taken as reference drugs over natural compounds for comparison and drug-likeness studies.

Chapter 14 discusses in brief the available pathways (both metabolic and non metabolic) that help in the synthesis of some major metabolites in bacterial population responsible for flourishing *ager*-flora in a natural way.

Chapter 15 focuses on the general steps to be followed in the *in silico* characterization of plant secondary metabolites, starting from literature mining, virtual screening, structural characterization, and structure-based drug designing.

Chapter 16 provides a brief overview on the databases and resources available to conduct *in silico* analysis of plant secondary metabolites and future prospects in utilizing the derived information to improve metabolite function and production in crops.

Bahadurgarh, Haryana, India — Devendra Kumar Choudhary
Ranchi, Jharkhand, India — Manoj Kumar
Noida, Uttar Pradesh, India — Ram Prasad
Noida, Uttar Pradesh, India — Vivek Kumar

Contents

About the Editors

Devendra K. Choudhary has over 16 years of research experience in Microbiology and Microbial Biotechnology and is currently working as Professor in the Department of Microbiology, Faculty of Life Sciences at PDM University, Bahadurgarh, Haryana. Before PDM he worked as an Associate Professor in Amity University, Noida. Before joining Amity University he spent several years in Mody University, Lakshmangarh, Rajasthan, as an Assistant Professor preceded by People's University and Barkatullah University, Bhopal, as an Assistant Professor cum Scientist.

Dr. Choudhary received his Ph.D. in Microbiology in 2005 from GB Pant University of Agriculture & Technology, Pantnagar, after having received his M.Sc. in Microbiology from MDS University, Ajmer, and qualifying CSIR-UGC-NET in 2002.

Dr. Choudhary has worked on GOI sponsored major projects as Principal Investigator (PI). Recently, he worked on two major projects for the DBT and SERB at Amity University preceded by the DST FAST-TRACK project at the Department of Biotechnology, Barkatullah University, Bhopal.

As an active researcher, Dr. Choudhary has published 80 research and review articles along with several book chapters for reputed journals and edited books. He is a recipient of the Indian National Science Academy (INSA) visiting and summer research fellowship 2014. He has been selected for prestigious Membership of the National Academy of Sciences, India.

Further, one of his Ph.D. students received the 2013 Dr. RS Rana Memorial best research award, sponsored by the Association of Microbiologists of India. Under Dr. Choudhary's supervision, his scientific team assigned two accession numbers from MTCC, IMTECH for submitted bacterial cultures (MTCC, 12057 & 12058) along with one MCC no. 2607. Most recently, his team has filed three patents with the India Patent Office, New Delhi, and others are in progress.

Manoj Kumar is an associate professor at the Centre for Life Sciences, Central University of Jharkhand. Dr. Kumar received his PhD in Plant Biotechnology from the prestigious Jawaharlal Nehru University, before being awarded consecutive postdoctoral fellowships by the IISc Bangalore and University of Pretoria. He currently serves as a reviewer for several journals, including *Phytoremediation*, *Journal of Soil Sediments*, and many more. Dr. Kumar's chief research interest is in the integration of microbial genetics with a breadth of plant physiological approaches to enable novel gene discovery.

Ram Prasad is senior assistant professor at the Amity Institute of Microbial Technology, Amity University, Uttar Pradesh, India. His research interests include plant-microbe interactions, sustainable agriculture, and microbial nanobiotechnology. Dr. Prasad has more than hundred publications to his credit, including research papers and book chapters, has edited or authored several books, and holds five patents (issued or pending). He has 11 years of teaching experience and was honored with the Young Scientist Award (2007) and Prof. J.S. Datta Munshi Gold Medal (2009) by the International Society for Ecological Communications; FSAB fellowship (2010) by the Society for Applied Biotechnology; Outstanding Scientist Award (2015) in the field of Microbiology by Venus International Foundation; and the American Cancer Society UICC International Fellowship for Beginning Investigators (USA, 2014). In 2014–2015, Dr. Prasad served as a Visiting Assistant Professor at the Department of Mechanical Engineering, Johns Hopkins University, USA.

Vivek Kumar is an associate professor at Amity University, Noida, Uttar Pradesh, India. He obtained his masters and doctoral degree from CCS Haryana Agricultural University, Hisar, Haryana, India. He currently serves on the editorial boards of many prominent international journals, e.g., *Environment Asia*, *International Journal of Biological and Chemical Sciences*, *Journal of Advanced Botany and Zoology*, and *Journal of Ecobiotechnology*. He is also a reviewer for the *Journal of Hazardous Materials*, *Science International*, *Acta Physiologiae Plantarum*, *Environment Science and Pollution Research*, and *Rhizosphere*. He has published 61 research papers, 19 book chapters, 6 review articles, and 2 books. Dr. Kumar also served as a Microbiologist for 8 years at the Department of Soil and Water Research, Public Authority of Agricultural Affairs and Fish Resources, Kuwait.

Chapter 1
Genes to Metabolites and Metabolites to Genes Approaches to Predict Biosynthetic Pathways in Microbes for Natural Product Discovery

Dharmesh Harwani, Jyotsna Begani, and Jyoti Lakhani

Abstract Microbes are exploited as a synthetic platform in the system biology where biochemistry of various biosynthetic pathways can be redesigned. De novo pathway biosynthesis is used to produce extremely valuable, high molecular weight compounds from renewable sources. The tools of bioinformatics and various other analytical approaches have played a significant role in in silico identification and characterization of novel chemical scaffolds of microbial origin by genome mining. However, the majority of the computational tools employ forward approach to link "genes" to their corresponding "metabolites," while the automated computational tools based on the retrosynthetic approach, to connect chemical structures of "metabolites" to their biosynthetic "genes," are still in their infancy. The retrobiosynthesis approach is an ingenious pathway design concept that has gained interest in the recent days because of its potential to assist in the redesigning of novel metabolic routes in a given biosynthetic pathway to produce the optimum levels of the targeted secondary metabolite. The approach employs a stepwise, backward to forward search to identify the most advantageous reactions to formulate and optimize the custom-made pathway evolution, whereas the retrosynthetic approaches are the computer-assisted framework to analyze various enzymatic reactions of the natural product biosynthetic pathways to connect and predict their corresponding genetic cluster. The main aim of the present chapter is to provide a brief overview of the retrobiosynthesis methodology and summarize about the available genome mining tools utilized in the forward and retrosynthetic approaches to envisage secondary metabolite biosynthetic gene clusters and vice versa as well as pathway design analysis and evolution. The perspectives on the further development of the genome-scale metabolic models and various other tools and databases in the field are also discussed.

D. Harwani (✉) · J. Begani
Department of Microbiology, Maharaja Ganga Singh University, Bikaner, Rajasthan, India
e-mail: dharmesh@mgsubikaner.ac.in

J. Lakhani
Department of Computer Science, Maharaja Ganga Singh University, Bikaner, Rajasthan, India

D. K. Choudhary et al. (eds.), *In Silico Approach for Sustainable Agriculture*,
https://doi.org/10.1007/978-981-13-0347-0_1

Keywords Genome mining · Secondary metabolites · Gene clusters · Retrobiosynthesis · Computational tools · Genome databases

1.1 Introduction

The natural products from microorganisms are the main source of drugs and enzymes and provide a huge variety of structural templates for the drug discovery and development. Chemical synthesis provides an alternate source to these natural products using which numerous compounds of therapeutic and industrial use have also been produced. De novo synthesis of these chemically complex compounds, in spite of their complicated natural makeup, is one of the most exciting and challenging discoveries (Fig. 1.1). These complex natural products always require a multistage synthesis step to complete an active compound. The methodology used for the synthetic designs is highly crucial for convergence and efficient synthesis of the natural products. The structures of the secondary metabolites from natural sources and their genetic circuits can help in the redesigning of the synthetic pathway to produce new chemical scaffolds (Walsh and Fischbach 2010; Medema and Fischbach 2015). The redesigning of their synthetic routes can be done using

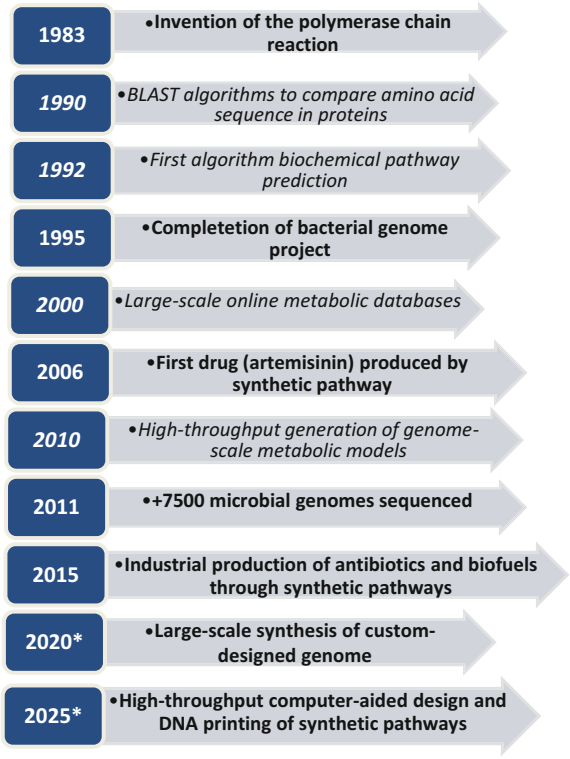


Fig. 1.1 Time scale of the main events for the evolution of de novo synthesis and pathway engineering. The evolution based on computational analysis is highlighted in italics, and the wet-lab-based evolution is shown in *bold letters*. *Future developments

"genes to metabolites" (forward) and "metabolites to genes" (retrosynthetic) approaches (Bachmann 2010; Cacho et al. 2015). The forward approach uses, the genetic information to deduce the final composition of the secondary metabolite, whereas, in contrast to the forward approach, the retrosynthetic approach begins with a known metabolite, and then attempts are made to determine the genes involved in the synthesis (Corey and Cheng 1989; Irschik et al. 2010).

The Nobel Prize winner, Professor Prof. E.J. Corey, was the first to describe the retrobiosynthetic analysis (synthon disconnection), which has become the most popular approach today for redesigning the chemical synthesis (Corey et al. 1961; Corey and Wipke 1969; Corey 1967, 1971; Corey and Cheng 1989; Corey and Guzman-Perez 1998; Corey and Link 1992). The retrosynthesis proceeds in a stepwise manner to synthesize the target molecule by repeated bond disconnections in reverse order until the starting material has not been achieved. In this article, we present an overview of the retrobiosynthesis method in general and in particular about the most common in silico tools of genome mining for the forward analysis of gene clusters responsible for the synthesis of secondary metabolites. The genome mining tools for the retrosynthetic analysis to connect the known secondary metabolites with their corresponding biosynthetic gene clusters are also described.

1.2 Why Microbial Secondary Metabolites Are Needed?

The emergence of multidrug resistance in pathogens has become a global issue and poses a major challenge to human health. A lot of efforts are being made worldwide to search for the alternate methods of treatments or discover novel and effective antimicrobials. According to the Health Ministry (UK government), antimicrobial resistance will lead to the ten million or more annual deaths by the year 2050 (Taylor et al. 2014). In this scenario, the quick solution is urgently needed to solve the issue of antibiotic resistance. It is important to quote here that the natural products are not only used as the antimicrobial drugs but are the main source of many lead compounds for other therapeutic applications. These are utilized as pharmacologically active compounds (antitumor, anti-inflammatory, neuroactive, antidepressants, anti-Alzheimer, cardioactive, platelet aggregation inhibitors, antioxidants, hypotensive, vasodilatory agents, nerve growth factor, interleukin and endothelin antagonists, estrogens, anti-ulcer, anti-allergic, antihistamine, anabolics, anesthetics, anticoagulants, hemolytics, hypocholesterolemics, immunoactive, immunosuppressants, immunomodulators, immunostimulators), enzyme inhibitors (peptidases, proteinases, glycosidases, amylases, HIV integrases, protein kinases, acetyl-coenzyme A acyltransferase, phosphatases, squalene synthetases, 3-hydroxy3-methylglutaryl-coenzyme A reductases, beta-lactamases, and monoamine oxidases), pesticides, and other activities (anti-parasitic, antimicrobial, herbicide, phytotoxic, plant growth regulatory, insecticide, nematocide, miticide, larvicide, anthelminthic, acaricide, ichthyotoxic, algicide, amoebicide antimetabolites, calcium antagonists, chelate-forming agents, siderophores, morphogenic agents, signaling and quorum-sensing

compounds, scavengers, bio-surfactants, feed additives, microbial hormones, and preservatives) (Bentley 1997; Berdy 2012). However, due to the re-isolation of known molecules and the lack of the innovative screening approaches, the pharmaceutical companies have reduced their research efforts for the discovery of natural products over the past two decades (Fischbach and Walsh 2009). Moreover, it is important to note that many compounds are synthesized chemically, but the most approved small molecule compounds are still the natural products or their derivatives (Newman and Cragg 2012).

1.3 Genome-Based Mining Strategies

With the increasing number of new and inexpensive sequencing techniques and available genetic data, it is now well-recognized that there is a huge number of yet-to-be-identified genetic clusters for secondary metabolite production in microbes (Weber et al. 2015a). The chemical diversity of secondary metabolites characterized so far is also enormous in nature, but intriguingly, the biochemistry of many of them is highly conserved. Consequently, many enzyme families of different categories for these secondary metabolites can be used to predict unique biosynthetic metabolic pathways. For this purpose, genome mining tools can be used by following these two main steps. In the first step, the genes coding for the conserved enzymes and protein domains, involved in the secondary metabolism, are identified using rule-based approaches. In the second step, the defined sets of rules are used to connect the natural products to the presence of such moieties. The more complex rules can be used if the specific genes are found to be expressed in the close proximity. As defined rules are prerequisite, the approach is not able to predict a completely new biosynthetic pathway involving the complex enzymatic reactions. To avoid this limit, different, rule-independent methods have been developed. These tools rely on automated computational analysis or framework to predict the genetic clusters of biosynthetic pathways coding for secondary metabolites. In this way, the field of computational analysis has proven to be a dynamic and versatile field to provide a possible answer to many unresolved matters pertaining to the discovery of novel natural products.

1.4 Retrograde Pathway Evolution

Horowitz provided the idea of retrograde evolution in the beginning of 1945. By definition, the retrobiosynthesis involves gene duplication and stepwise evolution of the biosynthetic pathways in the reverse order to that of biological synthesis. It means that the last enzymatic reaction will be considered the first, followed by subsequent enzymatic reactions involved in the pathway in the reverse order to the first enzymatic reaction (Horowitz 1945). The main idea of this theory is that

intermediates in the pathway cannot be estimated for their benefits and do not confer an optimal advantage to the host. However, if the pathway only produces the final product, then the individualistic reactions may be considered, and conditions can be improvised further for the continued production of that particular metabolite product. The whole scenario can be understood by the following example. Consider an organism which is supposed to be a heterotrophic for a particular growth intermediate C. The consumption of C increases, as growth commences and delimits it. To cope up with this condition, a metabolic strain can be engineered, by involving a mutant enzyme which can synthesize this intermediate C in a higher amount from the available precursors A and B. Under conditions, when the intermediate C has been consumed, the organism with the above mutant enzyme will have a certain growth advantage over the others. Future generations of the organism will also get benefits from it, if it is genetically stable (mutants reverting back to the wild type would not be able to synthesize intermediate C from A and B precursors and will disappear from the population). In addition to this if the concentration of either A or B is getting exhausted in the biosynthetic pathway, then, by applying the retrobiosynthesis approach, another genetically manipulated enzyme can be redesigned, which will allow the synthesis of A or B from precursors G or H. By extending the role of mutated enzyme in the retrobiosynthesis, the process will continue until there is a plentiful production of the final product which is supposed to be produced by the primary metabolism. In this model, pathway intermediates are assumed to be available from the immediate vicinity. But the present model is presumed to be controversial due to the debate that the intermediates involved in the biosynthetic pathways are highly unstable, and excessive buildup of the pathway precursors is highly unlikely (Fani and Fondi 2009; Jensen 1976). Therefore, the retrograde pathway evolution may be limited and only relevant to the evolution of compounds which are directly produced from the available molecules in prebiotic earth (Miller and Urey 1959; Johnson et al. 2008).

1.5 Pathway Evolution Through Retrobiosynthesis and Retrosynthetic Designs

The above approach has served as an inspiration for de novo creation of nonnatural pathway biosynthesis. Enzyme engineering and directed evolution are dependent on the laboratory-based changes in the genetic circuits and diversification of enzymes functioning in the biosynthetic pathway. The rationale of the evolution of biosynthetic enzymes in natural conditions is the selective pressure on the organism for existence, while the laboratory-based procedures determine the production of a particular product in a higher amount as defined and controlled by a researcher. The preference to apply the evolutionary model between the forward evolution and retrograde evolution, to create a nonnatural pathway, is generally inclined to the model of forward evolution (Morowitz 1999). The process of forward evolution

suggests that the directed evolution will resume with the first catalytic enzyme in the pathway to the last enzyme in a stepwise manner. As generalized above, in a forward model of evolution, all intermediates of the pathway provide a selective advantage in the development of the host, but in a nonnatural pathway model, one has to organize the entire enzymatic steps cautiously to control the synthesis of intermediates to get the final product in desirable amount. Furthermore, precautions are also required to monitor each enzyme involved in the pathway to screen important steps for the highest enzyme activity. Bachmann (2010) also reported that the bioretrosynthesis strategy is a dynamic practical approach to design a nonnatural synthetic pathway, and the selection of an appropriate assay for estimating the enzymatic activity is highly indispensable. Intriguingly, the enzymatic screening to design assay at each step of the development is exceptionally simple. The decisive enzymatic assay for the essential retrobiosynthesis reactions may eventually be able to facilitate the establishment of pathway flux to increase the overall production. Analogous to retrobiosynthesis, in synthetic organic chemistry, retrosynthesis approach is used also for nonnatural chemical synthesis (Corey 1991; Bachmann 2010). Unlike retrosynthesis, in retrobiosynthesis, many genetically different enzymes from many different sources are collected and tested to check their suitability to utilize as a key enzyme for the pathway engineering. In addition, the initial substrates and all other intermediary molecules implicated in the pathway evolution must also be available beforehand.

1.6 Genome-Based Mining Tools for Genes to Metabolites (Forward) Approach

Before the automatic tools were available, the key for bioinformatics-based analysis of genomes was manual mining. To perform this task, amino acids as a query for various proteins were searched using BLAST (Basic Local Alignment Search Tool) or PSI (Position-Specific Iterative BLAST) (Altschul et al. 1997). If alignment belongs to the known protein family, then it was used to generate Hidden Markov Models (HMMs) which function in concert with the software HMMER (Eddy 2011), and in this way, the gene groups were determined by analyzing the upstream and downstream genes. Today, the manual mining has been replaced by the automated computational tools, but still, the manual approaches, which do not follow the generalized rules, are exceedingly useful for the analysis and identification of gene clusters. Manual genome mining approach can further be improved using MultiGeneBlast, which facilitates BLAST-based analysis of the genes and operons (Medema et al. 2013). BLAST- and HMMER-based detections of gene clusters function very well with the low false positives and are globally exploited for the analysis of many different classes of secondary metabolites. BAGEL is a freely available online tool for global mining to identify RiPPs: ribosomally synthesized and posttranslationally modified peptides (van Heel et al. 2013; de Jong et al. 2006,

2010). It provides information on the classification and availability of the databases for known RiPPs. For the identification of genetic clusters pertaining to polyketide synthase (PKS), nonribosomal peptide synthetase (NRPS), and hybrid PKS/NRPS, several tools are available on the Internet. ClustScan (Cluster Scanner) is a Java-based desktop application that provides access to PKS and NRPS genome cluster data available with the ClustScan database (Starcevic et al. 2008). NP.searcher (Natural Products Searcher) is also a web-based program which is focused to estimate the structure of polyketide antibiotics (Li et al. 2009). NaPDoS (Natural Product Domain Seeker) (Ziemert et al. 2012) relies on BLAST and HMMER to detect ketosynthase domain of PKS and condensation domain of NRPS in the genomic data sets, which are subsequently classified into the functional categories. GNP (Genes to Natural Product Discovery Platform/Genome search (Johnston et al. 2015a, b; Ibrahim et al. 2012; Skinnider et al. 2015a, b) and GNP/PRISM (Prediction Informatics for Secondary Metabolomes) (Skinnider et al. 2015a) are web-based tools for the prediction of putative NRPS. The tools carefully detect and associate the information available with the metabolomics platform iSNAP: informatic search algorithm for natural products to identify corresponding peaks in liquid chromatography. SMURF (Secondary Metabolite Unique Regions Finder) can detect genes in the fungal PKS and NRPS and terpenoid gene clusters (Khaldi et al. 2010). CLUSEAN (CLUster SEquence ANalyzer) can automate the analysis of large datasets using text instead of interactive web pages (Weber et al. 2009).

The above-described tools are specialized in the analysis of specific category of secondary metabolites, but notably, antiSMASH (antibiotics and secondary metabolites analysis shell) provides search for the 44 different secondary metabolite categories (Weber et al. 2015b; Blin et al. 2013; Medema et al. 2011). With a variety of functions such as identification of genetic clusters, detailed annotation of the modular PKS and NRPS domain structures, substrate predictions and comparative genomics tools to identify biosynthesizing building blocks, genome-scale metabolic modeling, analysis of lanthipeptide pathways (Blin et al. 2014), and prediction of analogous gene clusters in sequenced genome databases in MIBiG (Minimum Information about a Biosynthetic Gene cluster) (Medema et al. 2015) dataset, antiSMASH is the most comprehensive tool to date for microbial genome mining. The entire rule-based genome mining approaches discussed so far are able to identify biosynthetic gene clusters which are already known but are futile to identify synthetic pathways, which are either yet-to-be-identified or use nonhomologous enzymes. But there are multiple numbers of options that can attempt to identify these biosynthetic gene clusters, independent of the prescribed set of rules. The software ClusterFinder (Cimermancic et al. 2014), which has been implemented with antiSMASH, uses HMM-based approach to access the genome for the detection of secondary metabolic pathways. The EvoMining (evolutionary-driven genome mining) approach (Cruz-Morales et al. 2015) identifies the cluster of metabolic genes linked to isozymes, even if they are phylogenetically diverse. By scanning the genome for these types of isozymes, it is possible to identify the associated genetic cluster, in spite of the gene sequence, coding for them are conserved. In addition to the usual mining tools discussed above, SEARCHPKS (Yadav et al.

2003) or NRPS-PKS/SBSPKS (Anand et al. 2010; Ansari et al. 2004) provide automated predictions about the multi-modular PKS and NRPS domains. The PKS/NRPS web server, predictive blast server, and 2metDB (Secondary Metabolite DataBase) (Bachmann and Ravel 2009) provide estimates based on the BLAST analysis against the signature sequences (Challis et al. 2000).

1.7 Databases Dedicated to Secondary Metabolite Biosynthetic Gene Clusters

In order to enable cross-species comparisons, many databases focus primarily on various aspects of secondary metabolism. The ClustScan database (Diminic et al. 2013), DoBISCUIT (Database of BIosynthesis Clusters Curated and InTegrated) (Ichikawa et al. 2013), and ClusterMine360 (Conway and Boddy 2013) have a limited collection of mostly hand-curated NRPS gene clusters. The r-CSDB (recombinant ClustScan database) (Starcevic et al. 2008) includes more than 20,000 in silico recombined sequences, which were expected to add new metabolic molecules into existing repository. Recently, standard MIBiG has been developed which has more than 1000 characterized biosynthetic gene clusters (Medema et al. 2015). In addition to these, databases are also based on many extensive gene sequencing efforts. There is a large collection of data in IMG-ABC (Integrated Microbial Genomes Atlas of Biosynthetic Gene Clusters) which is based on manually curated biosynthetic gene clusters, but they also contain automated mined data, sequencing of which was done at the US Department of Energy, JGI (Joint Genome Institute) (Hadjithomas et al. 2015). At present, the genomic data is collected from the culturable organisms. The fact, that only a fraction of these microbes can be cultivated (Harwani 2013), the unculturable microbes remain a large reservoir of unexploited genomes that may code for useful metabolites of therapeutic use. eSNAPD (environmental Surveyor of Natural Product Diversity) (Owen et al. 2013; Charlop-Powers et al. 2014, 2015) is a prominent system to detect synthetic gene clusters. Because the network can also use metadata, the data can be analyzed for a given sequence as well as for the location information of the sampling sites. Commonly known public databases, such as PubChem, PUBlic repository for information on CHEMical substances (Bolton et al. 2008); ChEMBL, a large-scale bioactivity database for drug discovery (Bento et al. 2014; Gaulton et al. 2012); and ChEBI, Chemical Entities of Biological Interest (Hastings et al. 2013; Degtyarenko et al. 2008), include information and annotations for a wide-ranging chemical compounds. In addition, commercial databases such as antiBASE, the natural compound identifier (Wiley-VCH, Weinheim, Germany), and the dictionary of natural products (Taylor and Francis Group LLC, USA) are also made available. Recently, several other databases have also been developed that are freely available or explicitly licensed. KNApSAcK, comprehensive species-metabolite relationship database website (Nakamura et al. 2014; Afendi et al. 2012), provides details about

various secondary metabolites in relation to their basic chemical properties and biological activities. Even though the KNApSAcK is dedicated to focus mainly on plant metabolites, but it also contains the details about microbial active compounds. A resource for natural compounds isolated from StreptomeDB (*Streptomyces* species DataBase) (Lucas et al. 2013; Klementz et al. 2016) is a database focusing on different secondary metabolic molecules from *Streptomycetes*. NORINE, a database of nonribosomal peptides (Caboche et al. 2008; Flissi et al. 2015), is the first database that is completely dedicated to provide data on NRPS, while a new web-accessible Bactibase, database for bacteriocin characterization (Hammami et al. 2007, 2010), contains data based on ribosomally synthesized antimicrobial peptides. Nowadays, lots of reports are published concerning about liquid chromatography-mass spectrometry (LC-MS) and nuclear magnetic resonance (NMR)-based metabolomics in natural product research (Wu et al. 2015a, b). However, a long history has been published for the chemical databases pertaining to the natural products and their related biosynthetic pathways, but different computational approaches have recently been linked to the chemical synthesis and classification of secondary metabolites. For the identification of large complex patterns of RiPPs and NRPS, a few specialized software programs have been found particularly more successful such as Pep2Path, automated mass spectrometry-guided genome mining of peptidic natural products (Medema et al. 2014); RiPPquest, a tandem mass spectrometry database search tool for identification of microbial ribosomally synthesized and posttranslationally modified peptides (Mohimani et al. 2014a); NRPquest, coupling mass spectrometry and genome mining for nonribosomal peptide discovery (Mohimani et al. 2014b); and Cycloquest, identification of cyclopeptides via database search of their mass spectra against genome databases (Mohimani et al. 2011).

1.8 Genome-Based Mining Tools for Metabolites to Genes (Retrosynthetic) Approach

The details of the biochemical conversions and various steps in the synthesis of secondary metabolites can be linked to the genes coding for their catalytic domains of enzymes in a given biosynthetic pathway by using computational methods. Many secondary metabolites have been characterized from microbes on the genetic basis, but, still, many genes coding for the biosynthesis of known natural compounds are yet to be identified (Horbach et al. 2009). In these microbes, for their yet-to-be characterized secondary metabolite gene clusters, genome mining tools based on retrosynthetic approach can be used. The main advantage of using retrosynthetic approach is that in spite of not having information on the genetic cluster connecting it to the secondary metabolite synthesis, it reliably helps in getting useful insights for the same. A variety of powerful computational algorithms are now accessible to decipher the metabolic steps involved in a biosynthetic pathway to connect two

different metabolites (McClymont and Soyer 2013; Soh and Hatzimanikatis 2010; Chou et al. 2009; Fenner et al. 2008). The retrosynthetic approach is also well documented to connect different metabolic routes of secondary metabolite synthesis in the plants (Moriya et al. 2010). In addition to this, various in silico automated tools are also in use to give a detailed overview of the biological degradation pathway analysis (Gao et al. 2011; Martin et al. 2009). The computational tools like ReBIT (Reserve Bank Information Technology) (Martin et al. 2009), from metabolite to metabolite (FMM) (Chou et al. 2009), and PathPred (Pathway Prediction Server) (Moriya et al. 2010) have also been developed to study the biological synthesis of new secondary metabolites using synthetic biology. PathPred focuses on deciphering pathways of secondary metabolites from plants, whereas the main focus of the majority of computational tools utilized in retrosynthetic analysis lies primarily on the investigation of primary metabolites and degradation pathways. The information about these metabolic routes is well documented in KEGG (Kyoto Encyclopedia of Genes and Genomes) database (Kanehisa 2002, 2016), whereas the information on the biological synthesis of natural products is available in various other scientific reports. PathPred predicts biosynthetic pathways involved in the degradation of xenobiotic compounds by using a RPAIR, a reactant-pair database, representing chemical changes in enzymatic reaction database containing biochemical patterns of transformation of substrate products (Kotera et al. 2004). Pep2Path assists in linking tandem mass spectra of nonribosomal peptides to their respective gene clusters by accepting MS-derived NRP mass shift sequence or an amino acid stretch or concerned gene sequence (Medema et al. 2014). The input in mass shifts is converted into amino acid tag first, and then the gene sequence is scanned by employing antiSMASH for possible nonribosomal peptide synthetase gene clusters. The Bayesian algorithm is utilized in Pep2Path, in order to decipher the possibility of an amino acid in a query tag that is involved in the synthesis or not for the presumed NRPS, and thereafter, the final calculations are made for the complete gene cluster.

The biological synthesis of PK and NRP molecules engages a small number of reactions as compared to the other biosynthetic pathways; for this reason, they can be disposed to the retrosynthetic analysis for deciphering their biosynthetic gene clusters. These kinds of analyses can be accomplished using Smiles2Monomers, a tool to link chemical structures to biological structures (Dufresne et al. 2015), or GRAPE (Generalized Retrobiosynthetic Assembly Prediction Engine) (Dejong et al. 2016) tool. These tools can decode the possible monomeric structure of a metabolic compound from a given database to link them to their biosynthetic gene clusters. This kind of analysis is accomplished using the antiSMASH database and GARLIC (Global Alignment for Natural Products Cheminformatics) tool (Dejong et al. 2016). The algorithms used in the retrosynthetic analysis (Law et al. 2009) of natural compounds make use of backward search, using a defined set of rules of biochemical transformations. A computational framework known as BNICE (Biochemical Network Integrated Computational Explorer) (Hatzimanikatis et al. 2005) predicts enzymatic reactions by bond electron matrices (Ugi et al. 1979) using which several specific elementary transformations (Tipton and Boyce 2000; Leber et al. 2009) are presented to the KEGG database (Kanehisa et al. 2008). Similarly, an algorithm

"molecular signature descriptor" (Faulon et al. 2008) predicts catalytic enzymes and is capable enough to generate the structure of a secondary metabolic compound. Likewise, based on 50 reaction rules, another retrosynthetic frameworks have also been developed (Cho et al. 2010). DESHARKY, automatic design of metabolic pathways for optimal cell growth (Rodrigo et al. 2008), is a prediction tool which relies on searches for all the possible routes to connect the enzymatic reactions of a metabolic pathway to the target compound. Subsequently, the thermodynamic feasibility, as well as the energy cost associated with the transcription and translation processes, is also estimated. Cho group have developed a method that predicts the reactions in the database, categorized by its type, and ranks the pathway accordingly. The web server, RetroPath, retrosynthetic metabolic pathway design (Carbonell et al. 2011), employs stringent but a manual search of the catalytic reactions based on the molecular signatures of the compounds involved. The other methods that help to estimate the effects of a new biosynthetic pathway on the host are known as OptStrain, a computational framework for redesign of microbial production systems (Pharkya et al. 2004), which uses flux analysis to provide suggestive about the optimization of the production if subtle changes are made in the host gene expression.

1.9 Concluding Remarks

Connecting secondary metabolites to their corresponding gene clusters using retrosynthetic approach would not only expand our current understanding of the complex biological synthesis but will also assist in the discovery of "much needed" novel natural products. Similarly, in silico tools developed for the forward analysis to identify novel chemical scaffolds of therapeutic and other beneficial use have also played an imperative role in the history of natural product research and facilitated enormously in linking the known genetic clusters of the biosynthetic pathways to their corresponding metabolites. These computational tools utilize complex, sequence-based algorithms for their analysis and domain predictions. The knowledge about tailoring enzymes, modular patterns, and iterative use of catalytic signature domains will technically improve the analytical capacity of various computational tools described so far. Both forward (genes to metabolites) and retrosynthetic (metabolites to genes) approaches exclusively rely either on the available information of gene sequences or the structural features of catalytic domains and enzymatic reactions implicated in the pathways of secondary metabolite synthesis. Availability of a large number of genome sequences has opened up the future prospects for the integrative analysis of the entire biosynthetic pathways by deliberated sequence manipulations of various catalytic domains. The retrosynthetic analysis is an important strategy, whose impact on the natural product research is enormous, supremacy of which is ever increasing in chemistry, biology, and medicine. It is imperative to note and, of course, it is highly logical to consider that the natural products will remain the important biological source than the randomly

synthesized chemical compounds, and the Mother Nature has unearthed only a tiny fraction of its huge store of molecular treasure. Upon the discovery and exploration of yet-to-be-discovered natural metabolites, we will definitively be having solutions against the current crises of "no-drug-development" and growing multidrug-resistant pathogens. In addition to this, the various computational approaches developed to date have also contributed tremendously in the identification and characterization of many novel biosynthetic gene clusters coding for essential secondary metabolites. Still, many challenges are there and will certainly appear in the near future. But undoubtedly, the forward and retrosynthetic approaches assisted by the tools of genome mining will come up with the possible solutions. Today, these approaches have become an essential part of methodology in the drug discovery programs. These methods are able to significantly complement the older approaches and provide the unforeseen dimensions in the natural product research.

References

Afendi FM, Okada T, Yamazaki M, Hirai-Morita A, Nakamura Y, Nakamura K (2012) KNApSAcK family databases: integrated metabolite-plant species databases for multifaceted plant research. Plant Cell Physiol 53(2):e1

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Anand S, Prasad MV, Yadav G, Kumar N, Shehara J, Ansari MZ (2010) SBSPKS: structure based sequence analysis of polyketide synthases. Nucleic Acids Res 38:W487–W496

Ansari MZ, Yadav G, Gokhale RS, Mohanty D (2004) NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. Nucleic Acids Res 32:W405–W413

Bachmann BO (2010) Biosynthesis: is it time to go retro? Nat Chem Biol 6:390–393

Bachmann BO, Ravel J (2009) Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. Methods Enzymol 458:181–217

Bentley R (1997) Microbial secondary metabolites play important roles in medicine; prospects for discovery of new drugs. Perspect Biol Med 40:364–394

Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M (2014) The ChEMBL bioactivity database: an update. Nucleic Acids Res 42:D1083–D1090

Berdy J (2012) Thoughts and facts about antibiotics: where we are now and where we are heading. J Antibiot 65:385–395

Blin K, Medema MH, Kazempour D, Fischbach M, Breitling R, Takano E (2013) antiSMASH 2.0–a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Res 41:W204–W212

Blin K, Kazempour D, Wohlleben W, Weber T (2014) Improved lanthipeptide detection and prediction for antiSMASH. PLoS One 9:e89420

Bolton E, Wang Y, Thiessen PA, Bryant SH (2008) PubChem: integrated platform of small molecules and biological activities. In: Wheeler R, Spellmeyer D (eds) Annual reports in computational chemistry, vol 4. American Chemical Society, Washington, DC, pp 217–241

Caboche S, Pupin M, Leclere V, Fontaine A, Jacques P, Kucherov G (2008) NORINE: a database of nonribosomal peptides. Nucleic Acids Res 36:D326–D331

Cacho RA, Tang Y, Chooi YH (2015) Next-generation sequencing approach for connecting secondary metabolites to biosynthetic gene clusters in fungi. Front Microbiol 5:774

Carbonell P, Planson AG, Fichera D, Faulon JL (2011) A retrosynthetic biology approach to metabolic pathway design for therapeutic production. BMC Syst Biol 5:122
Challis GL, Ravel J, Townsend CA (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. Chem Biol 7:211–224
Charlop-Powers Z, Owen JG, Reddy BV, Ternei MA, Brady SF (2014) Chemical biogeographic survey of secondary metabolism in soil. Proc Natl Acad Sci U S A 111:3757–3762
Charlop-Powers Z, Owen JG, Reddy BV, Ternei MA, Guimaraes DO, de Frias UA (2015) Global biogeographic sampling of bacterial secondary metabolism. elife 4:e05048
Cho A, Yun H, Park JH (2010) Prediction of novel synthetic pathways for the production of desired chemicals. BMC Syst Biol 4:35
Chou CH, Chang WC, Chiu CM, Huang CC, Huang HD (2009) FMM: a web server for metabolic pathway reconstruction and comparative analysis. Nucleic Acids Res 37:W129–W134
Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 158:412–421
Conway KR, Boddy CN (2013) ClusterMine360: a database of microbial PKS/NRPS biosynthesis. Nucleic Acids Res 41:D402–D407
Corey EJ (1967) General methods for the construction of complex molecules. Pure Appl Chem 14 (1):19–38
Corey EJ (1971) Centenary lecture. Computer-assisted analysis of complex synthetic problems. Q Rev Chem Soc 25:455–482
Corey EJ (1991) The logic of chemical synthesis – multistep synthesis of complex natural carbogenic molecules. Angew Chem Int Ed Eng 30:455–465
Corey EJ, Cheng XM (1989) Logic of chemical synthesis. Wiley, New York
Corey EJ, Guzman-Perez A (1998) Angew Chem Int Ed Engl 37:389
Corey EJ, Link JO (1992) A general, catalytic, and enantioselective synthesis of .alpha.-amino acids. J Am Chem Soc 114(5):1906–1908
Corey EJ, Wipke WT (1969) Computer-assisted design of complex organic syntheses. Science 166:178–192
Corey EJ, Ohno M, Vatakenchery PA, Mitra RB (1961) Total synthesis of d, l-longifolene. J Am Chem Soc 83(5):1251–1253
Cruz-Morales P, Martínez-Guerrero CE, Morales-Escalante MA, Yáñez-Guerra LA, Kopp JF, Feldmann J (2015) Recapitulation of the evolution of biosynthetic gene clusters reveals hidden chemical diversity on bacterial genomes. bioRxiv. https://doi.org/10.1101/020503
de Jong A, van Hijum SA, Bijlsma JJ, Kok J, Kuipers OP (2006) BAGEL: a web-based bacteriocin genome mining tool. Nucleic Acids Res 34:W273–W279
de Jong A, van Heel AJ, Kok J, Kuipers OP (2010) BAGEL2: mining for bacteriocins in genomic data. Nucleic Acids Res 38:W647–W651
Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A (2008) ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res 36:D344–D350
Dejong CA, Chen GM, Li H (2016) Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching. Nat Chem Biol 12:1007–1014
Diminic J, Zucko J, Ruzic IT, Gacesa R, Hranueli D, Long PF (2013) Databases of the thiotemplate modular systems (CSDB) and their in silico recombinants (r-CSDB). J Ind Microbiol Biotechnol 40:653–659
Dufresne Y, Noe L, Leclere V (2015) Smiles 2 monomers: a link between chemical and biological structures for polymers. J Chem inform 7:62
Eddy SR (2011) Accelerated profile HMM searches. PLoS Comput Biol 7:e1002195
Fani R, Fondi M (2009) Origin and evolution of metabolic pathways. Phys Life Rev 6:23–52
Faulon JL, Misra M, Martin S, Sale K, Sapra R (2008) Genome scale enzyme metabolite and drug-target interaction predictions using the signature molecular descriptor. Bioinformatics 24 (2):225–233

Fenner K, Gao J, Kramer S, Ellis L, Wackett L (2008) Data-driven extraction of relative reasoning rules to limit combinatorial explosion in biodegradation pathway prediction. Bioinformatics 24:2079–2085

Fischbach MA, Walsh CT (2009) Antibiotics for emerging pathogens. Science 325:1089–1093

Flissi A, Dufresne Y, Michalik J, Tonon L, Janot S, Noe L (2015) Norine, the knowledgebase dedicated to non-ribosomal peptides, is now open to crowdsourcing. Nucleic Acids Res 44(D1): D1113–D1118

Gao J, Ellis LB, Wackett LP (2011) The University of Minnesota pathway prediction system: multi-level prediction and visualization. Nucleic Acids Res 39:W406–W411

Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A (2012) ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 40:D1100–D1107

Hadjithomas M, Chen IM, Chu K, Ratner A, Palaniappan K, Szeto E (2015) IMG-ABC: a knowledge base to fuel discovery of biosynthetic gene clusters and novel secondary metabolites. MBio 6:e00932

Hammami R, Zouhir A, Ben Hamida J, Fliss I (2007) BACTIBASE: a new web-accessible database for bacteriocin characterization. BMC Microbiol 7:89

Hammami R, Zouhir A, Le Lay C, Ben Hamida J, Fliss I (2010) BACTIBASE second release: a database and tool platform for bacteriocin characterization. BMC Microbiol 10:22

Harwani D (2013) The great plate count anomaly and the unculturable bacteria. Int J Sci Res 2:350–351

Hastings J, de Matos P, Dekker A, Ennis M, Harsha B, Kale N (2013) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for. Nucleic Acids Res 41:D456–D463

Hatzimanikatis V, Li C, Ionita JA, Henry CS, Jankowski MD, Broadbelt LJ (2005) Exploring the diversity of complex metabolic networks. Bioinformatics 21(8):1603–1609

Horbach R, Graf A, Weihmann F, Antelo L, Mathea S, Liermann JC (2009) Sfp-type 4-′-phosphopantetheinyl transferase is indispensable for fungal pathogenicity. Plant Cell 21:3379–3396

Horowitz NH (1945) On the evolution of biochemical syntheses. Proc Natl Acad Sci U S A 31:153–157

Ibrahim A, Yang L, Johnston C, Liu X, Ma B, Magarvey NA (2012) Dereplicating nonribosomal peptides using an informatic search algorithm for natural products (iSNAP) discovery. Proc Natl Acad Sci U S A 109:19196–19201

Ichikawa N, Sasagawa M, Yamamoto M, Komaki H, Yoshida Y, Yamazaki S (2013) DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. Nucleic Acids Res 41:D408–D414

Irschik H, Kopp M, Weissman KJ, Buntin K, Piel J, Muller R (2010) Analysis of the sorangicin gene cluster reinforces the utility of a combined phylogenetic/retrobiosynthetic analysis for deciphering natural product assembly by trans-AT PKS. Chembiochem 11:1840–1849

Jensen RA (1976) Enzyme recruitment in evolution of new function. Annu Rev Microbiol 30:409–425

Johnson AP, Cleaves HJ, Dworkin JP, Glavin D, Lazcano A, Bada JL (2008) The Miller volcanic spark discharge experiment. Science 322:404–404

Johnston CW, Connaty AD, Skinnider MA, Li Y, Grunwald A, Wyatt MA (2015a) Informatic search strategies to discover analogues and variants of natural product archetypes. J Ind Microbiol Biotechnol 43(2):293–298

Johnston CW, Skinnider MA, Wyatt MA, Li X, Ranieri MR, Yang L (2015b) An automated genomes-to-natural products platform (GNP) for the discovery of modular natural products. Nat Commun 6:8421

Kanehisa M (2002) The KEGG database. Nucleic Acids Res 30(1):42–46

Kanehisa M (2016) KEGG bioinformatics resource for plant genomics and metabolomics. Methods Mol Biol 1374:55–70

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y (2008) KEGG for linking genomes to life and the environment. Nucleic Acids Res 36. Database:D480–D484

Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH (2010) SMURF: genomic mapping of fungal secondary metabolite clusters. Fungal Genet Biol 47:736–741

Klementz D, Doring K, Lucas X, Telukunta KK, Erxleben A, Deubel D (2016) Streptome DB 2.0 – an extended resource of natural products produced by streptomycetes. Nucleic Acids Res 44 (D1):D509–D514

Kotera M, Hattori M, Oh HM, Yamamoto R, Komeno T, Yabuzaki J (2004) RPAIR: a reactant-pair database representing chemical changes in enzymatic reactions. Genome Inform 15:P062

Law J, Zsoldos Z, Simon A, Reid D, Liu Y, Khew SY, Johnson AP, Major S, Wade RA, Ando HY (2009) Route designer: a retrosynthetic analysis tool utilizing automated retrosynthetic rule generation. J Chem Inf Model 49(3):593–602

Leber M, Egelhofer V, Schomburg I, Schomburg D (2009) Automatic assignment of reaction operators to enzymatic reactions. Bioinformatics 25(23):3135–3142

Li MH, Ung PM, Zajkowski J, Garneau-Tsodikova S, Sherman DH (2009) Automated genome mining for natural products. BMC Bioinformatics 10:185

Lucas X, Senger C, Erxleben A, Gruning BA, Doring K, Mosch J (2013) Streptome DB: a resource for natural compounds isolated from *Streptomyces* species. Nucleic Acids Res 41:D1130–D1136

Martin CH, Nielsen DR, Solomon KV, Prather KL (2009) Synthetic metabolism: engineering biology at the protein and pathway scales. Chem Biol 16:277–286

McClymont K, Soyer OS (2013) OSMetabolic tinker: an online tool for guiding the design of synthetic metabolic pathways. Nucleic Acids Res 41:e113

Medema MH, Fischbach MA (2015) Computational approaches to natural product discovery. Nat Chem Biol 11:639–648

Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res 39:W339–W346

Medema MH, Takano E, Breitling R (2013) Detecting sequence homology at the gene cluster level with MultiGeneBlast. Mol Biol Evol 30(5):1218–1223

Medema MH, Paalvast T, Nguyen DD, Melnik A, Dorrestein PC, Takano E, Breitling R (2014) Pep2Path: automated mass spectrometry-guided genome mining of peptidic natural products. PLoS Comput Biol 10(9):e1003822

Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K (2015) Minimum information about a biosynthetic gene cluster. Nat Chem Biol 11:625–631

Miller SL, Urey HC (1959) Organic compound synthesis on the primitive earth. Science 130 (3370):245–251

Mohimani H, Liu WT, Mylne JS, Poth AG, Colgrave ML, Tran D (2011) Cycloquest: identification of cyclopeptides via database search of their mass spectra against genome databases. J Proteome Res 10:4505–4512

Mohimani H, Kersten RD, Liu WT, Wang M, Purvine SO, Wu S (2014a) Automated genome mining of ribosomal peptide natural products. ACS Chem Biol. 2014 9:1545–1551

Mohimani H, Liu WT, Kersten RD, Moore BS, Dorrestein PC, Pevzner PA (2014b) NRPquest: coupling mass spectrometry and genome mining for nonribosomal peptide discovery. J Nat Prod 77:1902–1909

Moriya Y, Shigemizu D, Hattori M, Tokimatsu T, Kotera M, Goto S (2010) PathPred: an enzyme-catalyzed metabolic pathway prediction server. Nucleic Acids 38:W138–W143

Morowitz HJ (1999) A theory of biochemical organization, metabolic pathways, and evolution. Complexity 4:39–53

Nakamura Y, Afendi FM, Parvin AK, Ono N, Tanaka K, Hirai Morita A (2014) KNApSAcK metabolite activity database for retrieving the relationships between metabolites and biological activities. Plant Cell Physiol 55:e7

Newman DJ, Cragg GM (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. J Nat Prod 75:311–335

Owen JG, Reddy BV, Ternei MA, Charlop-Powers Z, Calle PY, Kim JH (2013) Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. Proc Natl Acad Sci U S A 110:11797–11802

Pharkya P, Burgard AP, Maranas CD (2004) OptStrain: a computational framework for redesign of microbial production systems. Genome Res 14:2367–2376

Rodrigo G, Carrera J, Prather KJ, Jaramillo A (2008) DESHARKY: automatic design of metabolic pathways for optimal cell growth. Bioinformatics 24:2554–2556

Skinnider MA, Dejong CA, Rees PN, Johnston CW, Li H, Webster AL (2015a) Genomes to natural products PRediction informatics for secondary metabolomes (PRISM). Nucleic Acids Res 43 (20):9645–9662

Skinnider MA, Johnston CW, Zvanych R, Magarvey NA (2015b) Automated identification of depsipeptide natural products by an informatic search algorithm. Chembiochem 16(2):223–227

Soh KC, Hatzimanikatis V (2010) DREAMS of metabolism. Trends Biotechnol 28:501–508

Starcevic A, Zucko J, Simunkovic J, Long PF, Cullum J, Hranueli D (2008) ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. Nucleic Acids Res 36:6882–6892

Taylor J, Hafner M, Yerushalmi E, Smith R, Bellasio J, Vardavas R, Bienkowska-Gibbs T, Rubin J (2014) Estimating the economic costs of antimicrobial resistance – model and results. RAND Europe, Cambridge

Tipton K, Boyce S (2000) History of the enzyme nomenclature system. Bioinformatics 16:34–40

Ugi I, Bauer J, Brandt J, Friederich J, Gasteiger J, Jochum C, Schubert W (1979) New applications of computers in chemistry. Angew Chem 18(2):111–123

van Heel AJ, de Jong A, Montalban-Lopez M, Kok J, Kuipers OP (2013) BAGEL3: automated identification of genes encoding bacteriocins and (non-) bactericidal post translationally modified peptides. Nucleic Acids Res 41:W448–W453

Walsh CT, Fischbach MA (2010) Natural products version 2.0: connecting genes to molecules. J Am Chem Soc 132:2469–2493

Weber T, Rausch C, Lopez P, Hoof I, Gaykova V, Huson DH (2009) CLUSEAN: a computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. J Biotechnol 140:13–17

Weber T, Charusanti P, Musiol-Kroll EM, Jiang X, Tong Y, Kim HU (2015a) Metabolic engineering of antibiotic factories: new tools for antibiotic production in actinomycetes. Trends Biotechnol 33:15–26

Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R (2015b) antiSMASH 3.0 – a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res 43:W237–W243

Wu C, Choi YH, van Wezel GP (2015a) Metabolic profiling as a tool for prioritizing antimicrobial compounds. J Ind Microbiol Biotechnol 43(2–3):299–312

Wu C, Kim HK, van Wezel GP, Choi YH (2015b) Metabolomics in the natural products field – a gateway to novel antibiotics. Drug Discov Today Technol 13:11–17

Yadav G, Gokhale RS, Mohanty D (2003) SEARCHPKS: a program for detection and analysis of polyketide synthase domains. Nucleic Acids Res 31:3654–3658

Ziemert N, Podell S, Penn K, Badger JH, Allen E, Jensen PR (2012) The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. PLoS One 7:e34064

Chapter 2
Role of Functional Bacterial Phylum Proteobacteria in *Glycine max* Growth Promotion Under Abiotic Stress: A Glimpse on Case Study

Anukool Vaishnav, Amrita Kasotia, and Devendra Kumar Choudhary

Abstract The PGPR elicit plant immunity referred to as induced systemic tolerance (IST) to cope up with abiotic stresses. The common modes of PGPR include fixing N_2, increasing the availability of nutrients in the rhizosphere, positively influencing root growth, promoting beneficial plant–microbe symbioses and succumb diseases. The present review deals case study of salt-tolerant rhizobacteria with respect to its functional plant growth promotional activities. Genomic DNA was isolated from bacterial strain AK-1, and gene-specific primers were used to amplify the 16S ribosomal DNA, ACC deaminase gene (*acdS* gene), IAA gene (*ipdC* gene), P-solubilizing gene (*gcd* and *gad* gene), ectoine production gene (*EctC* gene) and glycine betaine gene (*betA* gene). Gene amplification using specific gene primers showed sharp bands of the specific genes near to desired amplicon size. Bacterial-inoculated plants were exhibited superior tolerance against salt stress, as shown by their higher plant biomass, water content, chlorophyll content and lower osmotic stress injury as compared to non-inoculated plants during salt stress. Increased proline accumulation and antioxidant activity in bacterial-inoculated plants also contributed to salt tolerance. This study was conducted to assess the PGPR that are associated with the rhizosphere of soybean grown in semiarid areas of Rajasthan. We also sought to identify and characterize representative PGPR with respect to growth-promoting attributes and studied their salinity tolerance.

Keywords Abiotic stress · PGPR · Salt · Functional traits · *Pseudomonas*

A. Vaishnav · A. Kasotia
Amity Institute of Microbial Technology (AIMT), Amity University Campus, Gautam Buddha Nagar, Noida, UP, India

D. K. Choudhary (✉)
Department of Microbiology, Faculty of Life Sciences, PDM University, Bahadurgarh, Haryana, India
e-mail: dkchoudhary1@amity.edu; devmicro@rediffmail.com

D. K. Choudhary et al. (eds.), *In Silico Approach for Sustainable Agriculture*,
https://doi.org/10.1007/978-981-13-0347-0_2

2.1 Introduction

Abiotic stress is a very serious environmental factor worldwide that restricts the growth and productivity of plants worldwide. Once a seed germinates, plants being immobile are destined to stay where they are. Thus, they experience heavy selection pressure in their environments. Plants have developed many traits that help them to evolve and succeed across the globe under different environmental regimes. Following changes occurred in plants during stress condition like modification of cell wall, changes in cell cycle and cell division, production of osmolytes, etc. Many stress-responsive genes are also expressed, which includes the synthesis of osmoprotectants, detoxifying enzymes and transporters, as well as genes that encode regulatory proteins such as transcription factors, protein kinases and phosphatases during stress condition. Different plant species show different response with respect to their environment. Harsh environmental condition, which is harmful for one plant species (sensitive plant), might not be stressful for another plant (tolerant plant). Agriculture lands are frequently facing abiotic challenges like salinity, drought and high- and low-temperature conditions which ultimately affect plant growth and reduce crop yields, productivity and quality. Among these stresses, soil salinity contributes a major proportion in destruction of cultivated land area and reduction of crop productivity (Choudhary et al. 2016).

A saline soil is generally defined as one in which the electrical conductivity (EC) of the saturation extract (ECe) in the root zone exceeds 4 dSm^{-1} (approximately 40 mMNaCl). The yield of most plants is reduced at this salinity level. It has been estimated that the salinized areas are increasing at a rate of 10% annually which affected 20% of total cultivated and 33% of irrigated agricultural lands worldwide (Shrivastava and Kumar 2015). According to the Food and Agricultural Organization (FAO), if corrective measures are not taken, the arable land would be 50% salinized by the year 2050. Various number of salts, e.g. sodium chloride (NaCl), sodium sulphate (Na_2SO_4), sodium nitrate ($NaNO_3$), magnesium sulphate ($MgSO_4$), magnesium chloride ($MgCl_2$), potassium sulphate (K_2SO_4), calcium carbonate ($CaCO_3$), etc., could be dissolved in saline soil, although NaCl causes most of the salt problems for higher plants in nature. The dominant sources of soil salinity are weathering of parental rocks and rainfall. Weathering of rocks release various types of soluble salts, mainly chloride salts, while rainfall contains seawater salts, mainly sodium chloride. Rain containing 10 mg/kg of sodium chloride would affect the land by deposition of 10 kg/ha of salt during each 100 mm of rainfall per year. The other cause of accumulation is the intrusion of seawater on land which deposits a huge amount of salts in soils of coastal lands. Furthermore, human activities also contribute in soil salinity like poor-quality irrigation water, insufficient drainage, land clearing and the replacement of perennial vegetation with annual crops. An increase in the salts limit inhibits plant growth by an osmotic and ion stress. The former osmotic stress immediately comes over plant in accordance with a rise in salt levels outside the roots, which leads to inhibition of water uptake, cell expansion and lateral bud development (Shafique et al. 2014). The ionic stress develops when toxic level

of Na^+ accumulates in plants particularly in leaves over threshold level leading to leaf mortality with chlorosis and necrosis, a decreased essential cellular metabolic activities including photosynthesis and reduced enzyme activities. The harmful effects of salinity are not only on agricultural production but also on low economic returns due to high cost of cultivation (Munns and Tester 2008).

Increasing human population and reduction in land available for cultivation are two threats for agricultural sustainability. Today, it is big challenge to increase the efficiency and sustainability of global agriculture system. Because this system is regularly marked by scarcity of water resources, environmental pollution and increased salinization of soil and water. These challenges create continued poverty and food insecurity by which people become chronically malnourished.

2.2 Soybean *Numero Uno* Crop

Soybean (*Glycine max* L. Merrill) belongs to *Fabaceae* family, the subfamily *Faboideae* and the genus *Glycine*. The characteristic of this family member plants is that they have a mutualistic relationship with specific bacteria, in the case of soybean, *Bradyrhizobium japonicum*, a rhizobial species. During this interaction, the rhizobia provide nitrogen source to the plant by fixing atmospheric nitrogen. In turn, rhizobia receive sugars and minerals from the soybean plants to survive in the soil environment. It is an annual herbaceous bushy plant that can vary in height (0.2–2.0 m.). It grows well in warm and moist climate. A temperature of 26.5–30 °C appears to be optimum for most of the varieties. A well-drained and fertile loam soil with a pH between 6.0 and 7.5 are most suitable for cultivation. Its pods, leaves and stems are typically covered with brown or grey hairs. The leaves are trifoliate with three to four leaflets per leaf. Flowers are self-fertile white, pink or purple in colour and turn into pods. Each pod is 4–7 cm long, covered with hair and contains two to four seeds with 6–10 mm in diameter. The seed coat is hard, water resistant and protects the seed from damage and drying. Soybean seeds are an important source of oil and protein. Together, oil and protein content covered almost 60% of dry weight of soybean (protein at 40% and oil at 20%) and the remaining includes 35% carbohydrate and 5% ash. Its protein is used as a major source of dietary protein which contains all the essential amino acids particularly glycine, tryptophan and lysine, vitamins (A and D) and minerals used in place of cow's milk. The oil produced from soybean is highly digestible, contains no cholesterol and used mainly in cooking, margarine and salad dressings. Soybean is used in non-fermented foods (soy milk, tofu and tofu skin) and fermented foods (soy sauce, fermented bean paste, natto and tempeh). On the other hand, the extreme efforts on alternative sources of energy stimulated soy oil-based lubricant and fuel products that replace non-renewable petroleum products. Based on these attributes, soybean is the most promising component of the climate-smart agriculture concept (FAO 2013). It finds

a cheaper source of high-quality proteins and has potential to reduce malnutrition, a dominant problem in poor sections of society in the country.

According to USDA report (2015), soybean contributed about 60% of the total 536 million metric tons of oilseeds produced globally by major oil crops (sunflower, copra, peanut, cotton, palm and rapeseed).The USA and China were dominated countries in world soybean production through 1950–1970, growing more than 75 and 25% of the world soybean crop, respectively. In the early 1970s, a worldwide shortage of feed protein led to the initiation of soybean production in several other countries, most notably Argentina, Brazil and India. Comparing the world soybean production during 1965 and now, there has been a significant increase in 11.1 times. However, the USA and China shares of the world's soybean production had shrunk to 36% and 4%, respectively. In the present time, the top five countries, the USA (34%), Brazil (30%), Argentina (19%), China (4%) and India (3%), produce 90% of the world's soybeans (Fig. 2.1). It is exclusively grown in Northwest and Central part of India during kharif season (June to September). USDA (2015) estimated 9.0 million metric tons soybean production in all over India during 2014/2015 that was lowest in the last 5 years, while the area was estimated at 11.65 million hectares,

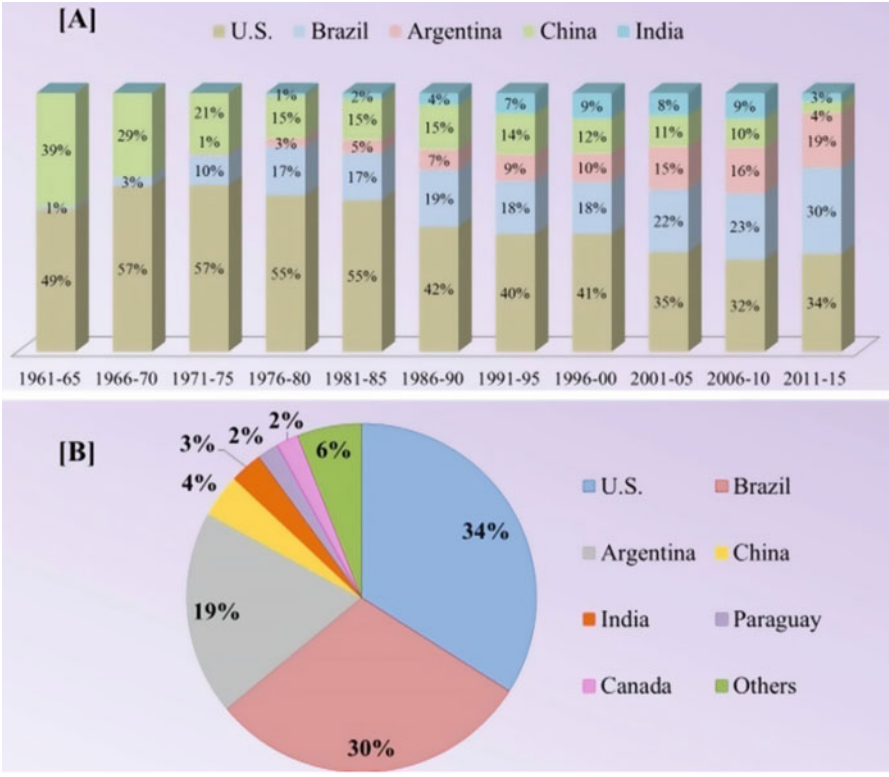


Fig. 2.1 World soybean production. (**a**) Soybean production in major countries during 1960–2015. (**b**) World soybean production during 2014–2015

increased by 0.7 million hectares from last year, and historically the second largest. These data indicate the poor levels of productivity of soybean crop in India as compared to other countries. The major soybean-producing states are Madhya Pradesh (60%), Maharashtra (30%) and Rajasthan (5%) (Fig. 2.2).Rajasthan consists of three climatic zones, namely, arid zone, semiarid temperate zone and semiarid tropical zone. Soybean is mainly grown in semiarid tropical zone of Rajasthan, regularly affected by high temperature, soil salinity, low pH and metal toxicity which cause a dramatic reduction in crop yield annually. The Rajasthan's soybean production is gradually decreasing since the last 5 years, and the lowest growing area and production were estimated 6.8 lakh Hectares and 5.6 lakh MT, respectively, in 2014/2015 (SOPA report 2014/2015) (Fig. 2.3).

Several abiotic factors are responsible for deprived production of soybean in India. Most of the areas under soybean cultivation are rainfed; hence, drought is a

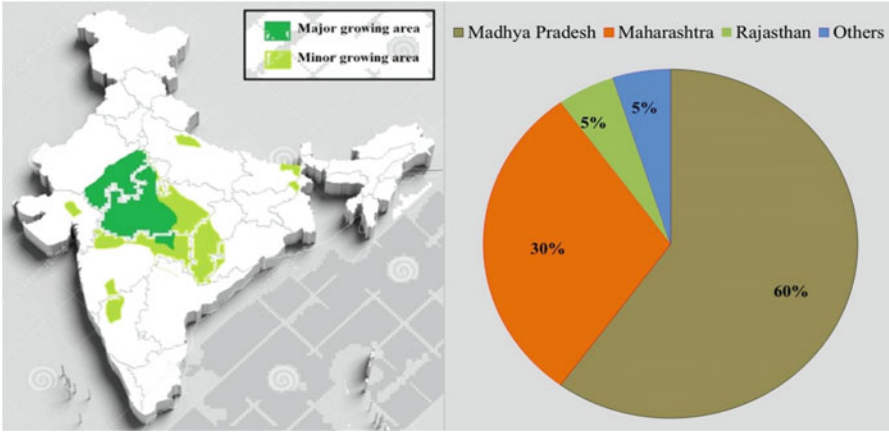


Fig. 2.2 Soybean production in different states of India

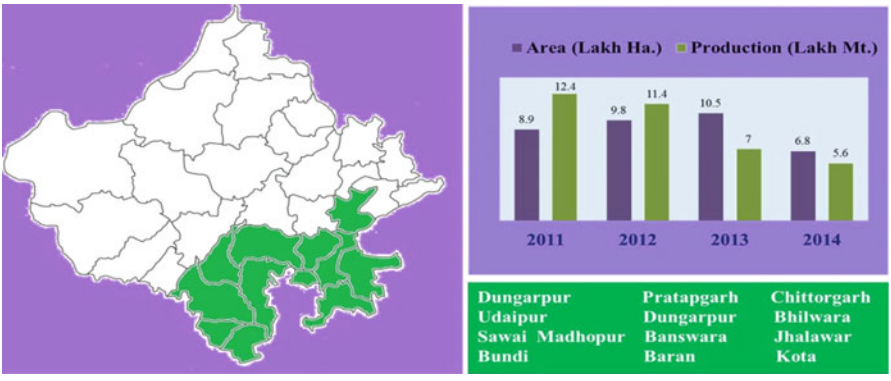


Fig. 2.3 Soybean production in Rajasthan

major constraint in soybean production. The Northwest areas of India are frequently facing an erratic behaviour of monsoon, which affects planting. A short intense drought can cause significant damage to plant and harm crop productivity. There is cross talk between drought and salt stress as both of these stresses create osmotic imbalance interference with nutrient availability and photosynthetic capacity and dehydration of the cell, which affects cell division, cell enlargement and differentiation (Nakashima et al. 2014). Soybeans consume large amounts of potassium (K) resulting low availability of K in soils. Potassium deficiency tends to have weak stems, which become more susceptible to pathogen attack.

Soybean is classified as moderately salt-sensitive crop, severely affected by decreasing symbiotic interaction which results in reduction of nodule formation and the amount of nitrogen fixed (Prudent et al. 2015). During salinity stress, malate concentration decreased that led to reduced bacteroid respiration and antioxidant content. High level of salts initiates water stress which leads to decrease in leghaemoglobin content, and accumulation of ureides in nodule decreases nitrogen fixation. Salinity is also reported to hasten the seed filling rate and decrease grain filling duration, which affect to seed protein and oil contents. Furthermore, the presence of large amount of salt alters the pH of soil, which leads to reduction in mobility of nutrient elements, such as P, K, Fe, Zn and Ca.

2.2.1 Selection and Characterization of PGPR Strains

A potent PGPR strain is selected from several root-colonizing bacteria by screening on the basis of their ability to produce PGP activity, inhibit the growth of various phytopathogens and a positive interaction with the host plant (Bhattacharyya and Jha 2012). Pure cultures of PGPR strains are applied on seeds in in vitro glasshouse trials. Seeds are treated with pure and fresh bacterial suspension and then planted in soil for test. During the experiment, those PGPRs that found significant enhances in plant growth and alleviates negative symptoms of stresses are selected for further field trials (Compant et al. 2005).Primary characterizations of new isolates are done based on biochemical characteristics as in *Bergey's Manual of Determinative Bacteriology* (Holt et al. 1994). After that, for molecular characterization, DNA- and RNA-based homology testing, ribosomal protein profiling through MALDI and fatty acid profiling through GC-MS analysis are considered as the most reliable tools for identification of PGPR strains (Bhattacharyya and Jha 2012). It is reported that 16S ribosomal RNA is a component of the 30S small subunit of prokaryotic ribosome. Through evolution this region of the gene remained conserved and hence widely used to define molecular phylogeny and taxonomy of bacteria since the last decade (Sun et al. 2008). PGPRs are becoming a frequent practice for enhancement of legume plants growth and its nutritional value. Most of the studies focused on co-inoculations of PGPR and rhizobium for the improvement of nitrogen-fixing

Table 2.1 PGPR-mediated IST against different stresses in soybean plant

Experimentations performed by	(PGPRs)	Description
Fernández-Bidondo et al. (2011)	*Paenibacillus rhizosphaerae*, *Glomus intraradices*	Enhanced dry biomass and plant growth
Stefan et al. (2011)	*Bacillus pumilus*	Enhanced nodulation and protein content
Salavati et al. (2012)	*Bradyrhizobium japonicum*	Proteomic analysis of early response of root
Wang et al. (2012)	*B. japonicum*	Lipo-chitooligosaccharides signal compounds induced stress-related genes under suboptimal temperature
Juge et al. (2012)	*B. japonicum, Azospirillum* and *Arbuscular mycorrhizae* (consortia experiment)	Enhanced proline, chlorophyll and antioxidant activity
Marks et al. (2013)	*B. diazoefficiens* and *Rhizobium tropici* (metabolites study)	Improved *Bradyrhizobium* spp. interaction along with enhanced growth and yields
Prudent et al. (2015)	*B. thuringiensis*, *B. japonicum* (consortia experiment)	Enhanced plant growth under drought stress
Algar et al. (2014)	Microbe associated molecular patterns (MAMPs)	Induced systemic resistance against *Xanthomonas axonopodis pv. glycines*
Ramesh et al. (2014a)	*B. aryabhattai*	Increased exchangeable and bound zinc in soil and finally assimilation in plants
Ramesh et al. (2014b)	*B. aryabhattai*	Increased plant growth parameters and P content
Masciarelli et al. (2014)	*B. amyloliquefaciens* and *B. japonicum* (consortia experiment)	Enhanced the capacity of colonization and increase the number of nodules
Armendariz et al. (2015)	*B. japonicum* and *Azospirillum brasilense*	Promoted plant growth under arsenic-contaminated soil
Simonetti et al. (2015)	*Pseudomonas fluorescens* and *B. subtilis*	Induced resistance against *Macrophomina phaseolina* pathogen attack

capacity under stress conditions (Egamberdieva et al. 2015). Some recent reports studied on PGPR effects on soybean are presented in Table 2.1.

Many resident microflora of stress environment perform all functions of life for survival of their own and associated biological entities. Some genera, like *Bacillus*, *Paenibacillus* and *Pseudomonas*, are actively being used to alleviate abiotic stresses (reviewed by Choudhary et al. 2016). In saline environment, halophilic bacteria and their metabolites have exhibited many potential which are suitable for vast agricultural, industrial and environmental applications. The successful restoration of plant growth under salinity condition after inoculation with halophilic bacteria provides the basis for a suitable alternative to improve crop growth and yield in saline soils. In this regard, the present study focused on application of salt-tolerant bacteria to

improve soybean plant growth and reduce negative effect of salt stress. Salt-tolerant bacterial isolates were recovered from soybean rhizospheric soil grown in semiarid region of Rajasthan, India. The mechanisms of bacterial-mediated IST in soybean plant were studied at biochemical and molecular level.

2.2.2 PGPR Characteristics: Case Studies

Study performed with reference strain AK (MTCC number 12058).

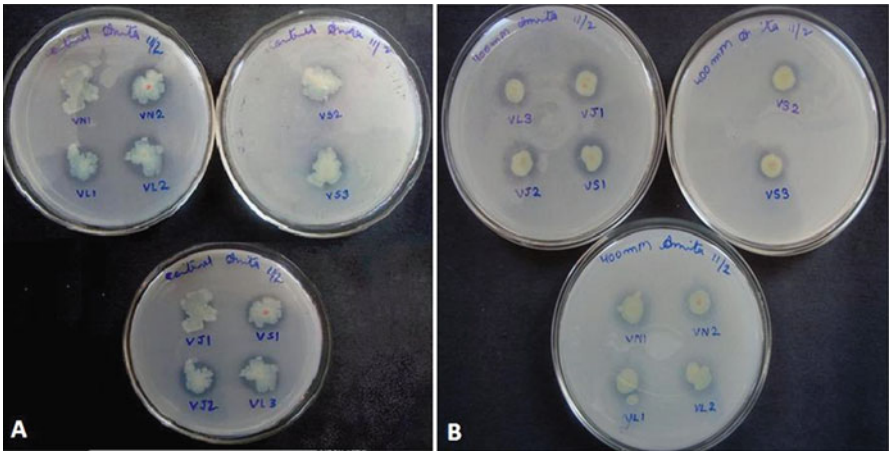


Fig. 2.4 Inorganic P solubilization. (**a**) 0 mMNaCl. (**b**) 400 mMNaCl by bacterial strains

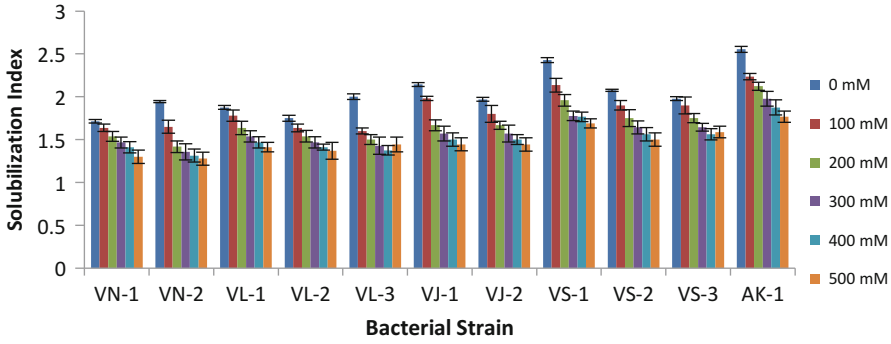


Fig. 2.5 Solubilization index bacterial strains in different salt concentrations. Values are means of 3 replications ± S.D

2.2.2.1 Qualitative and Quantitative Estimation of Phosphate Solubilization Under Salt

Inorganic phosphate solubilization in plate assay revealed that as the strains were able to solubilize phosphate in selected salt stress treatments 0–500 mMNaCl (Fig. 2.4a, b). All the recovered strains were able to solubilize Pi and form halo zone around the colonies. The strains AK-1 and VS-1 showed maximum solubilization index in salt stress as well as non-salt stress (Fig. 2.5).

The strain AK-1 showed solubilization index in the range of 2.55–1.77, and VS-1 showed in the range of 2.43–1.69. All the selected bacterial strains were checked for

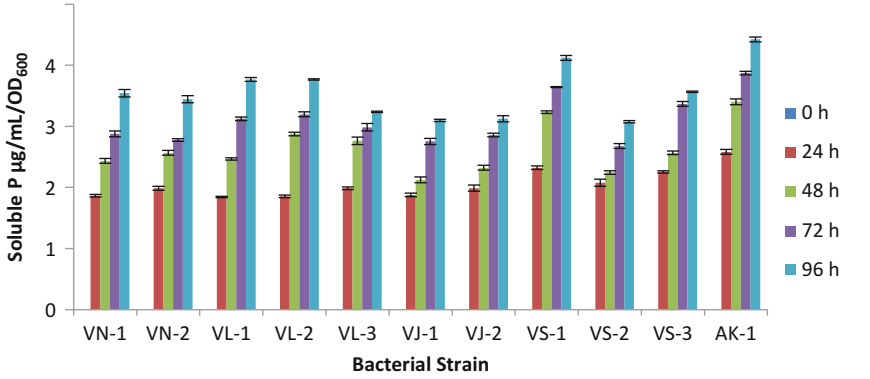


Fig. 2.6a Pi solubilization by bacterial strains in 0 mMNaCl. Values are means of 3 replications ± S.D.

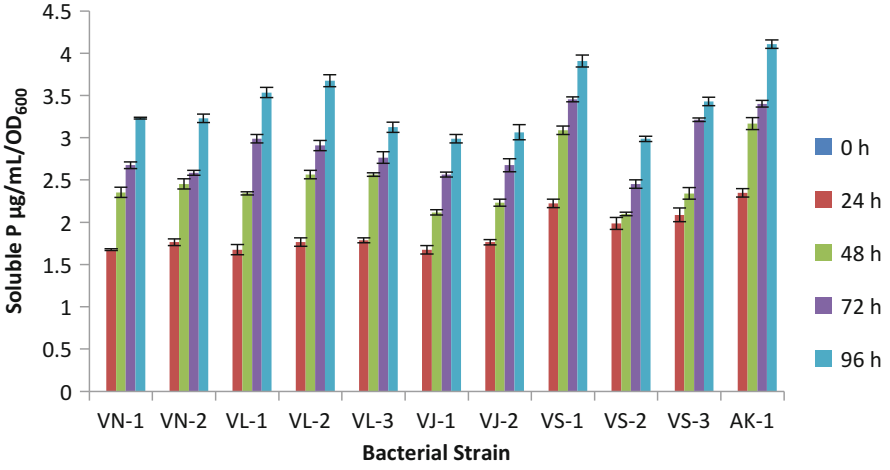


Fig. 2.6b Pi solubilization by bacterial strains in 200 mMNaCl. Values are means of 3 replications ± S.D

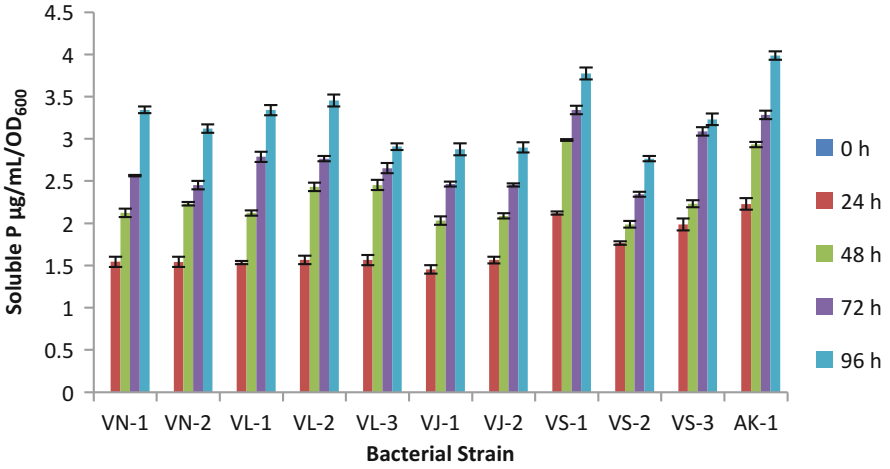


Fig. 2.6c Pi solubilization by bacterial strains in 300 mMNaCl. Values are means of 3 replications ± S.D

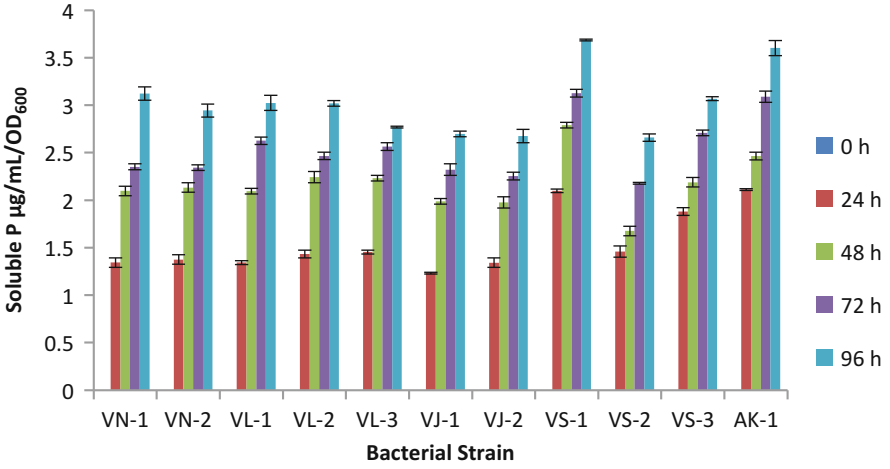


Fig. 2.6d Pi solubilization by bacterial strains in 400 mMNaCl. Values are means of 3 replications ± S.D

inorganic phosphate solubilization in different salt concentrations quantitatively also. All bacterial strains were able to solubilize inorganic phosphate in controlled condition. The tendency to solubilize Pi decreased as the concentrations of NaCl increased from 0 to 500 mM (Fig. 2.6a, 2.6b, 2.6c, 2.6d, and 2.6e). The two bacterial strains AK-1 and VS-1 were able to solubilize maximum Pi in all the salt concentrations. The strain AK-1 produced maximum soluble P of 4.43 µg/mL in 0 mMNaCl

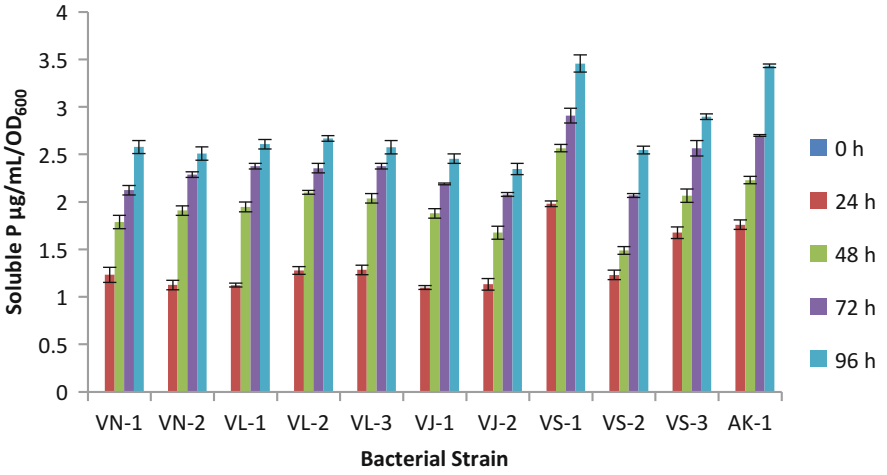
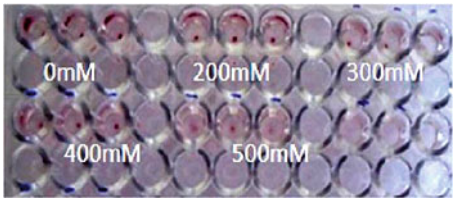


Fig. 2.6e Pi solubilization by bacterial strains in 500 mMNaCl. Values are means of 3 replications ± S.D

Fig. 2.7a ACC deaminase activity in microtiter plate assay by bacterial strain AK-1 in different salt concentrations



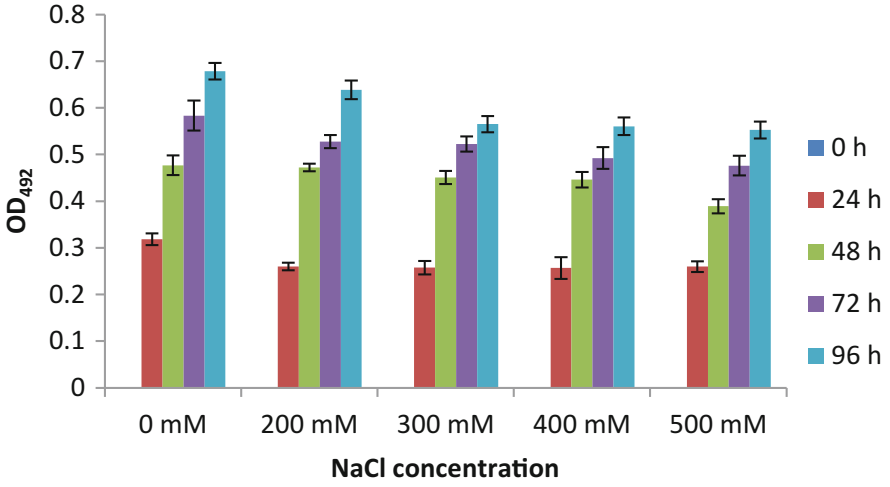


Fig. 2.7b ACC deaminase activity in microtiter plate assay by bacterial strain AK-1 in different salt concentrations. Values are means of 3 replications ± S.D

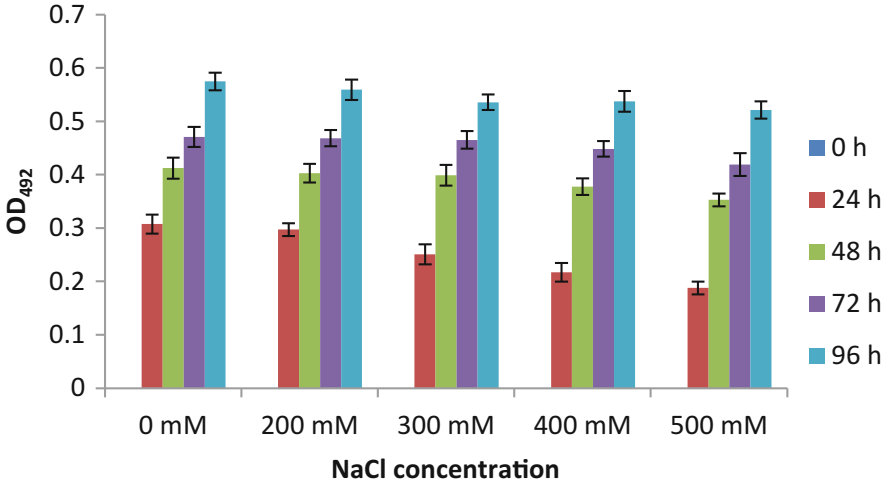


Fig. 2.7c ACC deaminase activity in microtiter plate assay by bacterial strain VS-1 in different salt concentrations. Values are means of 3 replications ± S.D

and 3.4 µg/mL in 500 mMNaCl, and the strain VS-1 produced maximum soluble P of 4.1 µg/mL NaCl and 3.4 µg/mL in 0 mM and 500 mMNaCl, respectively.

2.2.2.2 ACC Deaminase Activity in Different Salt Concentrations

Two bacterial strains AK-1 and VS-1 were tested for ACC utilization by bacterial strains as nitrogen source (Fig. 2.7a, 2.7b, and 2.7c). The strains were able to metabolize ACC and fulfil their nitrogen requirement. TTC acted as electron donor and respiratory indicator and formed pink-coloured insoluble compound with the growth of bacterial strains. As the bacterial strains were less stressed in less NaCl concentrations, so more coloured compound was formed in 0 mM than other NaCl concentrations (0–500 mM).

2.2.2.3 Proline Determination in Different Salt Concentrations

Compatible solute proline's concentration was determined in bacterial strain AK-1 which revealed that the bacterial strain was able to form proline as cellular osmolyte in salt stress (Fig. 2.8a). Cellular osmolyte tends to decrease with the increasing salt concentration from 200 to 500 mM due to decreased bacterial growth in salt stress. Maximum 300 µg/g weight proline accumulation was found in 200 mMNaCl (Fig. 2.8b). The control (0 mM) NaCl containing AK-1 showed minimum proline content of 160 µg/g weight proline.

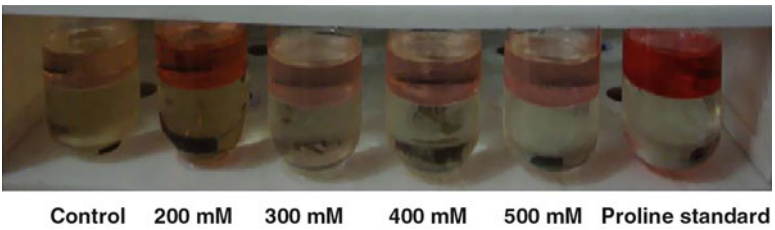


Fig. 2.8a Proline production in different salt concentration by bacterial strain AK-1

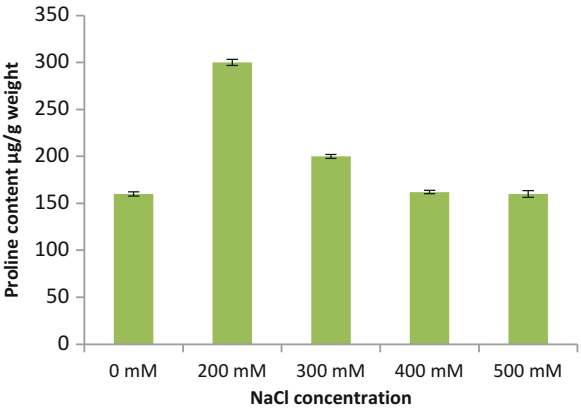


Fig. 2.8b Proline production in different salt concentration by bacterial strain AK-1. Values are means of 3 replications ± S.D

Table 2.2 Exopolysaccharide production by bacterial strain AK-1

Fresh weight (g mL^{-100})		Dry weight (g mL^{-100})	
0 mMNaCl	300 mMNaCl	0 mMNaCl	300 mMNaCl
0.65 ± 0.049	0.82 ± 0.054	0.32 ± 0.077	0.43 ± 0.079

Values are presented as means ± SD; $n = 5$

2.2.2.4 Exopolysaccharide Production

The beneficial role of exopolysaccharide-producing bacteria is known to remove various toxic heavy metals. Bacterial EPSs bind to cations including Na^+; when the population density of EPS-producing bacteria is increased in the root zone, it would decrease Na^+ available for the plant uptake and thus helps in alleviation of salinity stress (Khodair et al. 2008). Bacteria produced 0.65 g fresh weight of exopolysaccharide in absence of NaCl, but when the NaCl concentration increased to 300 mM, fresh weight increased to 0.82 g, likewise dry weight also increased to from 0.32 g in 0 mMNaCl to 0.43 g in 300 mMNaCl (Table 2.2). This increase in EPSs can be attributed to survival of bacteria in salinity stress which can contrarily

reduce available Na^+ for plant uptake in soil. EPSs are also known to bind soil particles closely and avoid forming drought stress.

2.2.2.5 Molecular Characterization of Genes in Bacteria

Genomic DNA was isolated from bacterial strain AK-1, and gene-specific primers were used to amplify the 16S ribosomal DNA, ACC deaminase gene (*acdS* gene), IAA gene (*ipdC* gene), P-solubilizing gene (*gcd* and *gad* gene), ectoine production gene (*EctC* gene) and glycine betaine gene (*betA*gene). Gene amplification using specific gene primers showed sharp bands of the specific genes near to desired amplicon size (Fig. 2.9a–g).

2.2.3 Sterilization and Viability and Seed Sensitivity

Seed sterilization was standardized by absence bacterial growth in nutrient both in minimum $HgCl_2$ concentration and treatment time. The seeds were found sterilized when treated with 0.1% $HgCl_2$ for 1 min (Fig. 2.10a). Absence of bacterial growth and seed germination on filter paper assay further standardized the seed sterilization protocol for plant experiments. Triphenyltetrazolium chloride is a redox indicator which was used as cellular respiration indicator to check the seed viability. TTC on reduction formed insoluble reddish-coloured formazan (Fig. 2.10b). The seeds selected for the study were found viable and showed reddish-coloured formazan.

Soybean seeds' salt susceptibility was performed on the six varieties, viz. PK 1050, JS 9560, NR 7, JS 7105, PK 1025 and JS 9305. Out of six varieties, JS 9560 was found most susceptible for 200 mMNaCl and least salt tolerant (Fig. 2.10c). Hence, soybean seed variety JS 9560 was selected for the plant experiments.

2.2.4 Standardization for Bacterial Inoculums

Soybean seed variety JS 9560 (salt sensitive) was inoculated with 1×10^8 CFU/mL bacterial densities for all the 11 bacterial strains in filter paper assay as shown in Fig. 2.11. All the 11 bacterial strains augmented root length and lateral roots up to some or greater extent (Table 2.3). The two bacterial strain AK-1 and VS-1 increased the root length and lateral root utmost nearly 17.5 cm and 28.8 by AK-1, 16.1 cm and 27.6 by VS-1 with 100% germination by both the bacterial inoculants. Hence, bacterial strain AK-1 was selected as bacterial inoculant for further studies on plants. Some of the experiments were performed on VS-1 also.

Study performed with reference strain AU (MTCC number 12057).

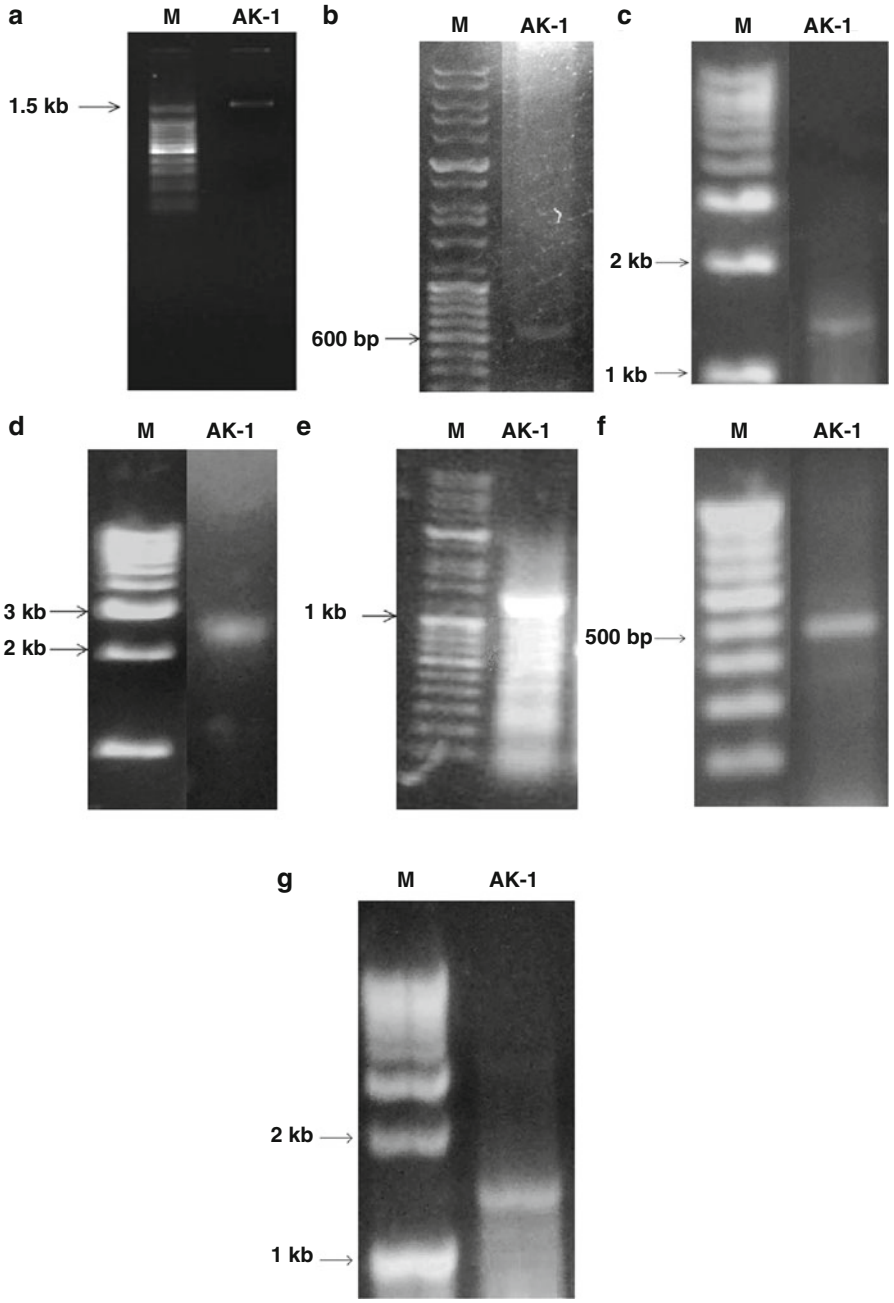


Fig. 2.9 PCR amplification of different genes of bacterial strain AK-1 on 1% agarose gel showing amplicons. M = marker DNA. (**a**) 16S ribosomal RNA of ~1.5 kb; (**b**) *acdS* gene ~650 bp; (**c**) *ipdC* gene ~1.4 kb; (**d**) *gcd* gene ~2.4 kb; (**e**) *gad* gene ~1.7 kb; (**f**) *EctC* gene ~500 bp; (**g**) *betA* gene ~1.5 kb

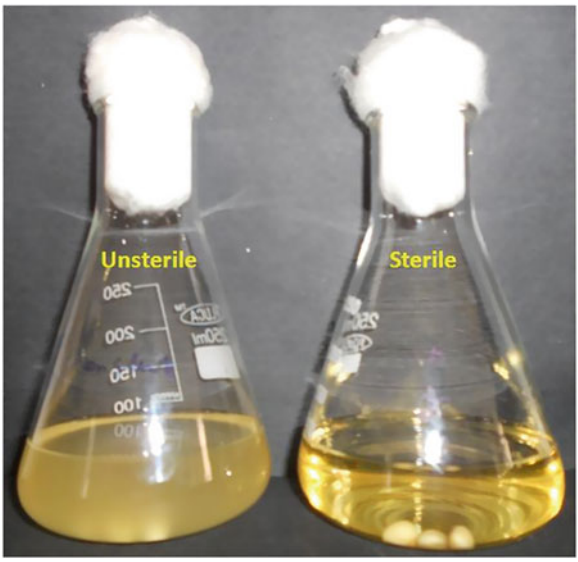


Fig. 2.10a Seed sterilization standardization

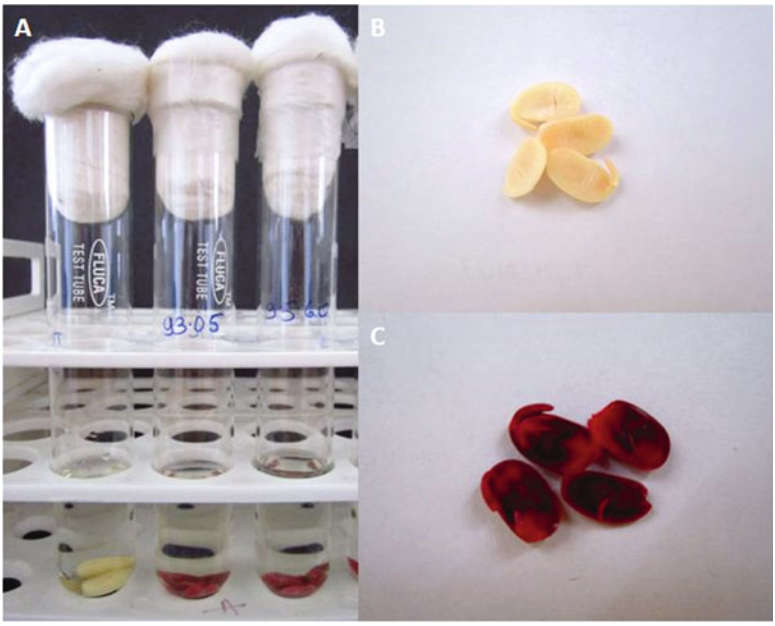


Fig. 2.10b Seed viability check for soybean. (**a**) In 1% TTC solution first tube showing control seed (0% TTC) and second and third tube showing the two seed varieties JS 9305 and JS 9560; (**b**) Control seeds; (**c**) viable seed showing reddish-pink-coloured formazan formation

Fig. 2.10c Salt-sensitive test for JS 9560 in 200 mMNaCl. Control with 0 mMNaCl and treated with 200 mMNaCl

Fig. 2.11 Bacterial inoculum standardization for the plant treatments

Table 2.3 Bacterial inoculum standardization

Bacterial strain	Germination %	Root length (cm)	Lateral roots (nos.)
Control	80	15.9 ± 1.92	18.6 ± 2.96
VN-1	80	13.5 ± 1.64	10 ± 3.03
VN-2	60	4 ± 0.81	3.6 ± 1.61
VL-1	80	9.3 ± 1.59	6.8 ± 1.44
VL-2	100	7.1 ± 0.79	8.4 ± 1.77
VL-3	20	2.7 ± 1.2	3.2 ± 1.43
VJ-1	80	11.14 ± 1.57	6.8 ± 1.34
VJ-2	60	8.6 ± 1.84	9.8 ± 2.72
VS-1	100	16.1 ± 1.43	27.6 ± 1.73
VS-2	80	10.34 ± 1.43	6.8 ± 1.82
VS-3	80	9.78 ± 1.13	13.6 ± 1.58
AK-1	100	17.5 ± 0.72	28.8 ± 1.33

Values are the means of 5 replications ± S.E

Table 2.4 Different bacterial strains growth under NaCl and PEG condition

S. No.	Strain	NaCl (%)	PEG (%)
1	AT	2–8	2–14
2	AT1	2–4	2–8
3	AT2	2–4	2–12
4	AT3	2–4	2–12
5	AT4	2–10	2–14
6	AT5	2–6	2–8
7	AT6	2–4	2–10
8	AT7	2–8	2–12
9	AT8	2–8	2–10
10	AT9	2–4	2–8
11	AT10	2–6	2–8
12	AT11	2–8	2–10
13	AT12	2–10	2–12
14	AT13	2–8	2–12
15	AM	2–8	2–12
16	AM1	2–4	2–10
17	AM2	2–4	2–8
18	AM3	2–10	2–12
19	AM4	2–4	2–8
20	AM5	2–12	2–14
21	AM6	2–4	2–8
22	AM7	2–4	2–8
23	AM8	2–6	2–10
24	AM9	2–4	2–8
25	AM10	2–4	2–10
26	AM11	2–6	2–10
27	AM12	2–8	2–8
28	AM13	2–6	2–8
29	AM14	2–6	2–10
30	AM15	2–6	2–10
31	AM16	2–4	2–10
32	AU	2–10	2–12
33	AU1	2–4	2–8
34	AU2	2–4	2–8
35	AU3	2–12	2–8
36	AU4	2–6	2–10
37	AU5	2–4	2–10
38	AU6	2–6	2–8
39	AU7	2–8	2–8
40	AU8	2–4	2–8
41	AU9	2–4	2–8
42	AU10	2–6	2–8
43	AU11	2–4	2–8

2.2.4.1 Bacterial Growth Under Stress Conditions

The results of the effect of salt and drought concentrations on rhizobacterial growth are presented in Table 2.4. The bacterial isolates differed in their ability to tolerate salinity in form of NaCl. All isolates could tolerate salinity up to 2% NaCl. A decrease in isolate number was observed with increase in NaCl concentration. A number of 19 isolates could tolerate 4% NaCl, 10 isolates could tolerate 6% NaCl, 8 isolates could tolerate 8% NaCl, 4 isolates could tolerate 10% NaCl, and 2 isolates were able to tolerate 12% NaCl concentration.

All isolates were screened for their ability to tolerate PEG wherein 3 isolates were able to tolerate maximum 14% PEG whereas 7, 13 and 20 isolates could tolerate 12%, 10% and 8% PEG, respectively. All isolates could tolerate drought up to 6% PEG.

2.2.5 PGP Traits

A total number of 14 bacterial isolates which were showed growth on 8%, 10% and 12% NaCl were selected for PGP traits characterization (Fig. 2.12). Two bacterial strains AT7 and AM were not able to produce any PGP properties. A number of three bacterial strains were able to produce only single one ACC deaminase activity. A total number of six bacterial strains were able to produce more than one PGP activity, and three bacterial strains (AU, AU7 and AT11) were producing all PGP properties tested (Table 2.5).

2.2.5.1 IAA Production on Different Salt Concentration

A total number of nine bacterial strains were able to produce IAA in the presence of 100 µg/ml L-tryptophan; however, four bacterial strains (AT, AT4, AM3 and AU)

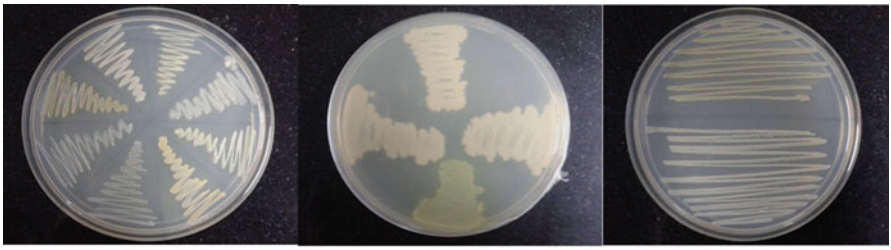


Fig. 2.12 Bacterial isolates showed growth on different NaCl concentration

Table 2.5 Characterization of selected salt-tolerant bacteria based on their PGP properties

S.No	Isolates	Plant growth promoting attributes					
		IAA	P-sol	Sid	ACC	EPS	VOCs
1	AT	+	+	+	–	+	–
2	AT4	+	–	–	–	+	–
3	AT7	–	–	–	–	–	–
4	AT8	–	–	–	+	–	–
5	AT11	+	+	+	+	+	+
6	AT12	+	+	–	–	+	–
7	AT13	–	–	–	+	–	–
8	AM	–	–	–	–	–	–
9	AM3	+	+	+	–	–	–
10	AM5	–	–	–	+	–	–
11	AM12	+	+	–	–	–	–
12	AU	+	+	+	+	+	+
13	AU3	+	–	–	–	+	–
14	AU7	+	+	+	+	+	+

+ and – indicates for positive and negative for specific PGP properties

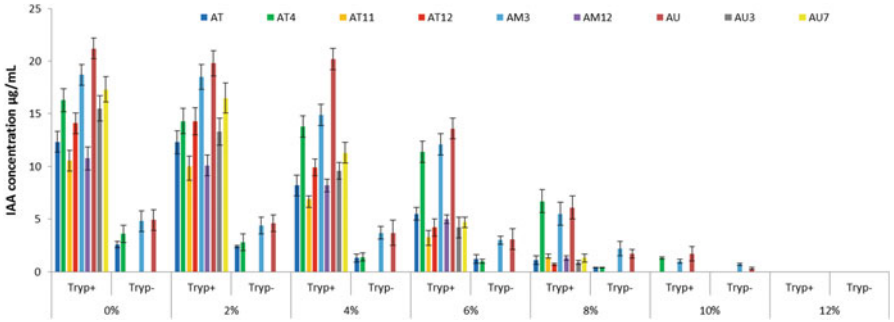


Fig. 2.13 Production of IAA for different bacterial isolates under NaCl condition. Values represent the means ± SD, $n = 3$

were able to produce IAA without L-tryptophan also. IAA production was found inversely proportional to salt concentrations, when the concentration of salt increased from 2% to 12% IAA production decreased. Only three bacterial isolates AU, AT4 and AM3 were able to retain sufficient IAA producing activity (>10 µg/mL) at 10% NaCl concentration, and other bacterial strains lost their activity after 8% NaCl concentration. During stress condition, the highest IAA production was observed in AU isolate (13.6 µg/mL) followed by AM3 (12.1 µg/mL) and AT4 (11.4 µg/mL) at 6% NaCl (Fig. 2.13).

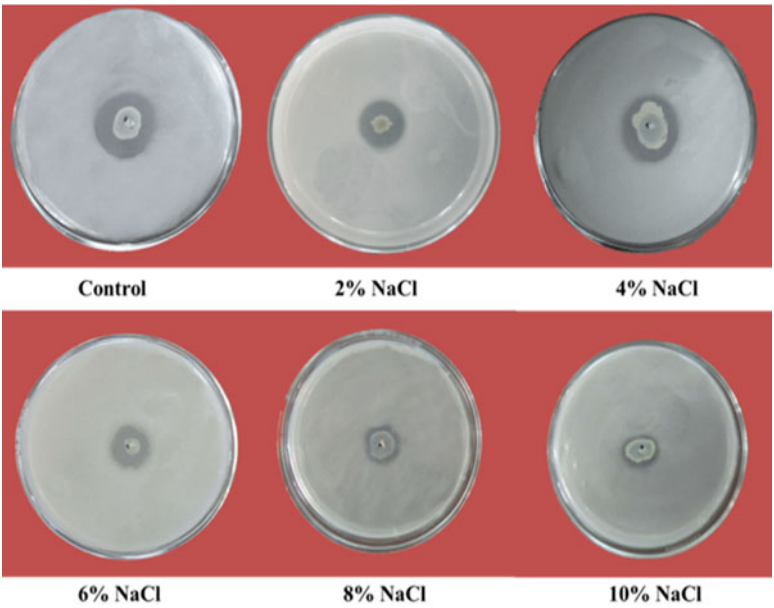


Fig. 2.14a Phosphate solubilization activity for AU bacterial isolate under different NaCl concentration

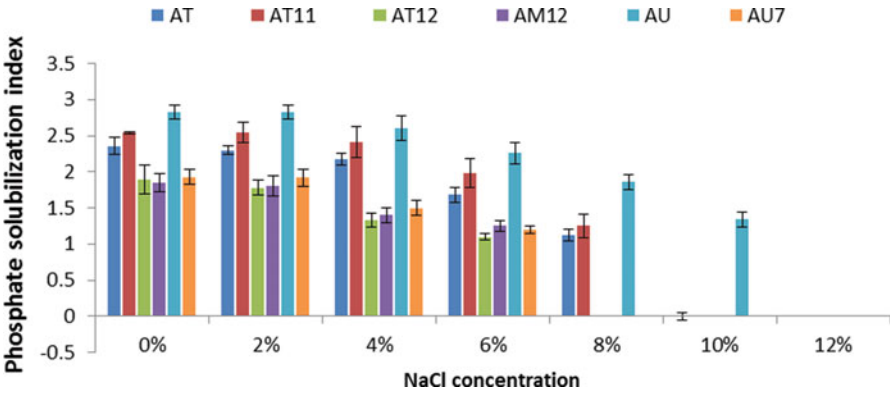


Fig. 2.14b Production of phosphate solubilization activity for different bacterial isolates under NaCl condition. Values represent the means ± SD, $n = 3$

2.2.5.2 Phosphate Solubilization on Different Salt Concentration

When the isolates were subjected to qualitative test for P solubilization using Pikovskaya agar plates, seven isolates were able to form clear halo around the colonies indicating positive results. The strain AU, AT and AT11 were showed

maximum solubilization index in salt stress as well as non-salt stress (Fig. 2.14a). The ranges of solubilization index were 2.83–1.86, 2.36–1.12 and 2.54–1.77 in AU, AT and AT11 isolates, respectively (Fig. 2.14b). Other isolates showed the development of hazy zones at higher salt concentration.

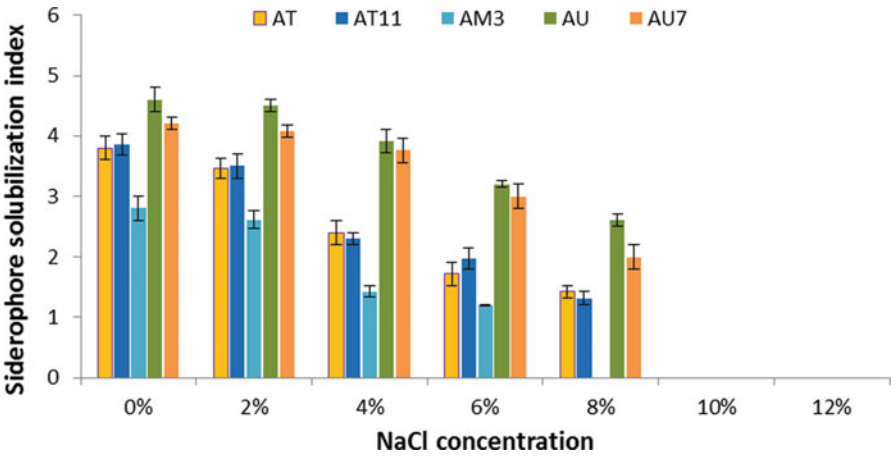


Fig. 2.15a Siderophore production activity for different bacterial isolates under NaCl condition. Values represent the means ± SD, $n = 3$

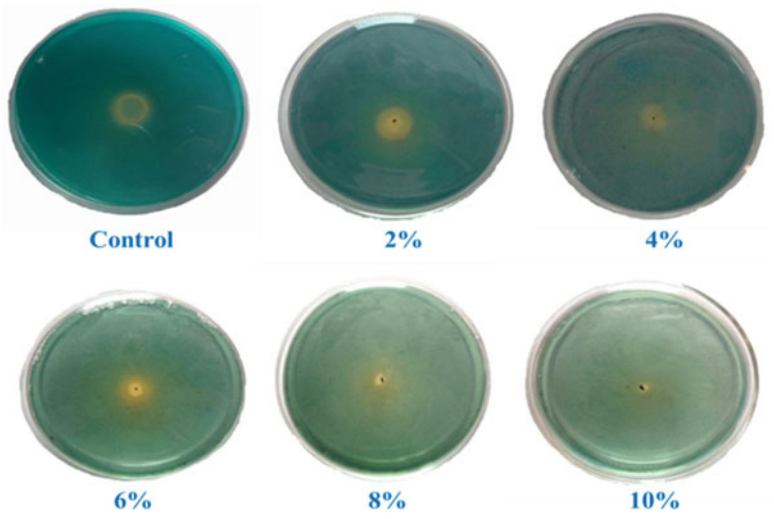


Fig. 2.15b Siderophore activity for AU bacterial isolate under different NaCl concentration

2.2.5.3 Siderophore Production on Different Salt Concentration

Qualitative estimation revealed that siderophore production was higher at 0% NaCl in a total number of five bacterial isolates as compared to other salt concentration. The index size was recorded in ranged from 4.6 to 1.2. Siderophore production decreased with increased salt concentration. The highest index during salt stress was measured in AU (2.6) followed by AU7 (2) at 8% as compared to other isolates (Fig. 2.15a and 2.15b).

2.2.5.4 ACC Deaminase Activity on Different Salt Concentration

The 14 strains were tested for ACC deaminase activity, and 6 strains were found to be positive. A number of three bacterial strains lost their activity at 6% NaCl; however, AU, AU7 and AT11 strains were able to retain their activity up to 8% NaCl. The highest activity was observed in cell-free extract of AU isolates at 0% NaCl (73 nmol α-ketobutyrate mg protein^{-1} h^{-1}) and 8% NaCl (29 nmol α-ketobutyrate mg protein^{-1} h^{-1}) (Fig. 2.16).

2.2.5.5 EPS Production on Different Salt Stress

A total number of seven bacteria were found to produce EPS production. The production was increased under salt stress. Bacterial isolates AT, AT11 and AT12 were showed highest production at 6%; AU7 and AU3 were showed at 8%; and AU was produced at 10% NaCl. The highest production was found in AT11 (1.3 g mL^{-100}) followed by AU (1.17 g mL^{-100}) (Fig. 2.17).

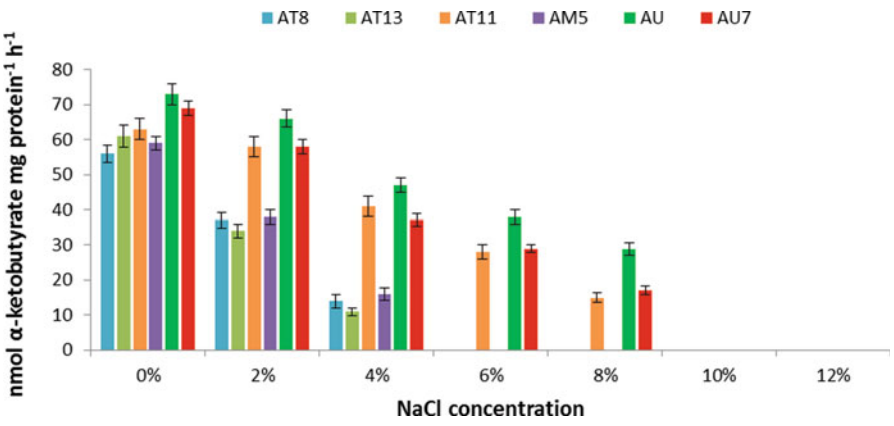


Fig. 2.16 ACC-deaminase activity for different bacterial isolates under NaCl conditions. Values represent the means ± SD, $n = 3$

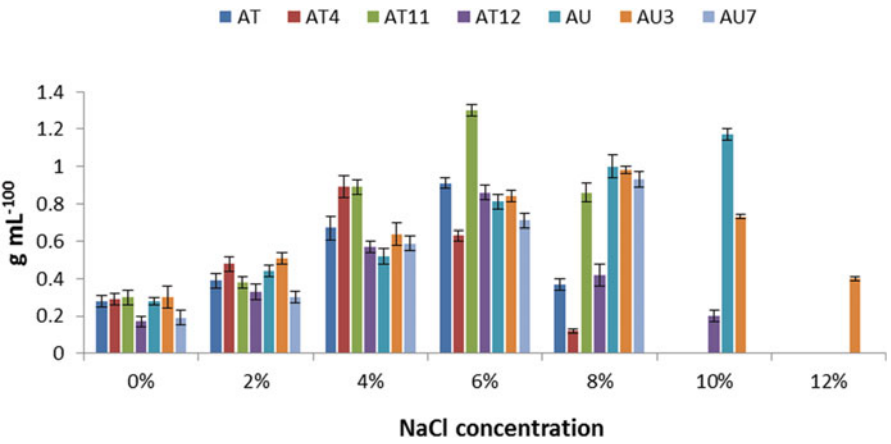


Fig. 2.17 Exopolysaccharide production assay for different isolates under NaCl conditions. Values represent the means ± SD, $n = 3$

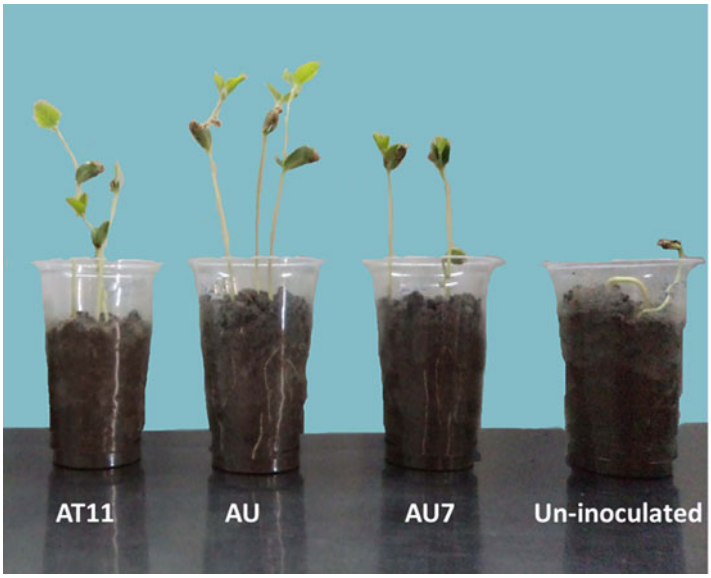


Fig. 2.18 Evaluation of plant growth promotion efficiency of three bacterial strains on soybean plant

2.2.5.6 Phylogeny Analysis of PGP Gene Properties

Gene amplification using specific gene primers showed sharp bands for *acds* (410 bp), *IaaM*(330 bp), *g6pd* (630 bp), sid (360 bp), tre (510 bp), nr (630 bp) and p5cs (250 bp) (Fig. 2.19). All genes were sequenced and submitted in NCBI;

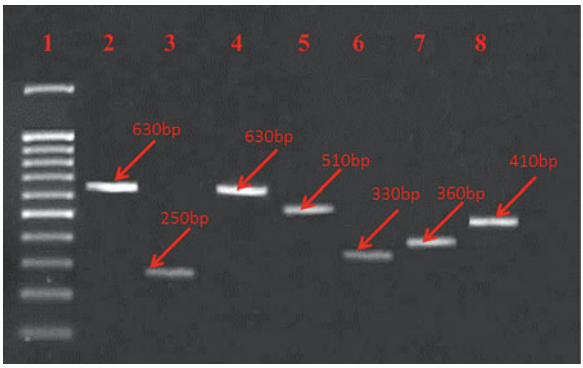


Fig. 2.19 PCR amplification of seven different PGP genes in AU bacterial isolate

Table 2.6 Details of gene sequences submitted in NCBI

NCBI accession number	Details	Bacterial strain
KU159726	ACC-deaminase (*acds*)	*Pseudomonas simiae*AU
KT805422	Nitrite reductase (*nr*)	*Pseudomonas simiae*AU
KT805423	Alpha amylase (*amy*)	*Pseudomonas simiae*AU
KT805424	Indole acetic acid (*IaaM*)	*Pseudomonas simiae*AU
KU204777	Siderophore (*sid*)	*Pseudomonas simiae*AU
KU204778	Gluconic acid (*g6pd*)	*Pseudomonas simiae*AU
KU204779	Proline (*p5cr*)	*Pseudomonas simiae*AU
KJ184311	16S ribosomal RNA gene	*Pseudomonas* sp. AU

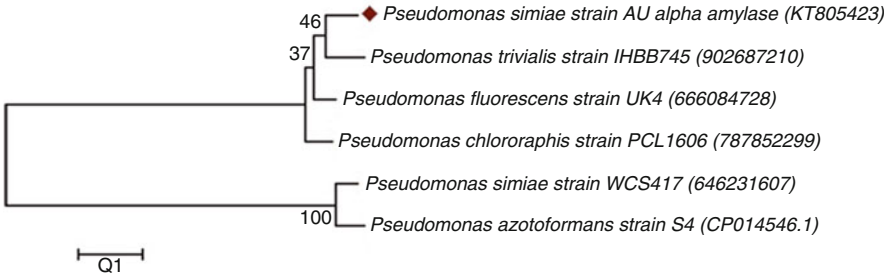


Fig. 2.20a Phylogenetic analysis of *acds* and *sid* gene of AU bacterial isolate

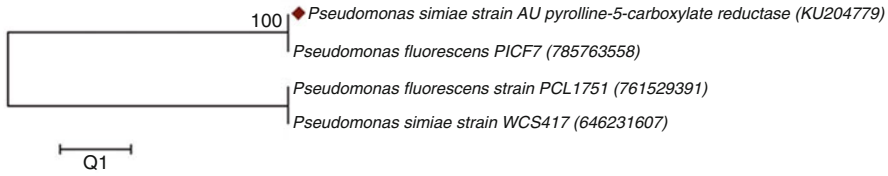


Fig. 2.20b Phylogenetic analysis of *p5cr* and *IaaM* isolated from AU bacterial isolate

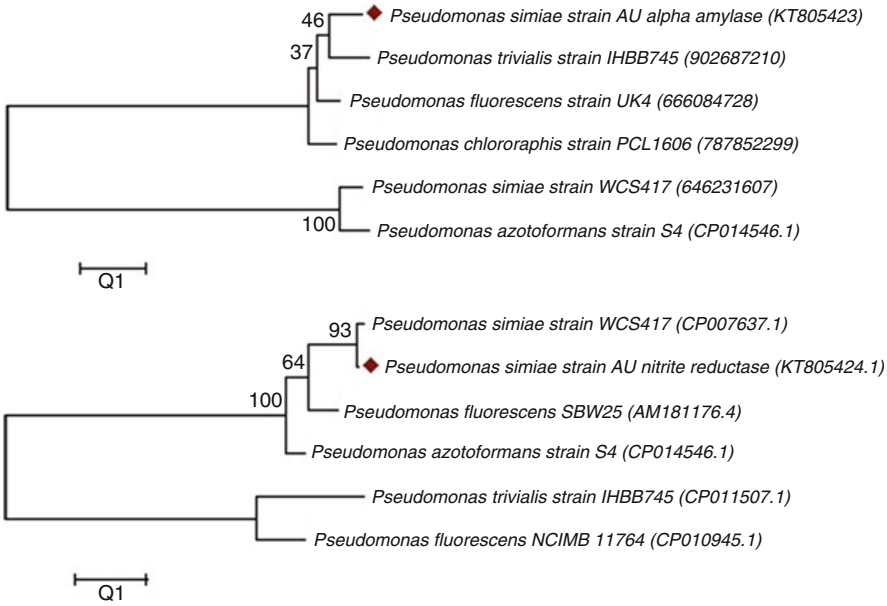


Fig. 2.20c Phylogenetic analysis of *amy* and *nr* isolated from AU bacterial isolate

GenBank accession numbers are listed in Table 2.6. As expected, all PGP genes of AU bacterial strain tested showed similar to the *P. fluorescence* intrageneric cluster. Regarding the phylogeny based on the *IaaM* and *nr*gene sequences, the AU bacterial strain forms an independent cluster, which includes *P. simiae* WCS417 (CP007637), whereas in *acds, g6pd, sid, tre* and *p5cs* genes phylogeny, AU bacterial strain was included with *P. fluorescens* PICF7 (CP005975) in an independent cluster. All other species from the genus *Pseudomonas* are found outside this cluster (Fig. 2.20a, 2.20b, and 2.20c).

It is consensus that salinity is a major limiting factor for crop productivity and a major cause of the reduction of cultivable lands in semiarid areas of the world. A wide range of strategies have been adopted to develop salt-tolerant crops but little success to date as few genetic traits for salt tolerance are identified (Schubert et al. 2009). Several studies have demonstrated that local adaptation of plants to habitat-imposed stresses is driven by their closely associated microbes (reviewed by Choudhary 2012). Soybean is a major oilseed crop in India, and it covers almost 4% of world's soybean production. It is one of the salt-sensitive (glycophytes) agricultural crops. Salinity detrimentally affects growth and development of soybean by severely reducing the metabolic processes such as CO_2 assimilation, oil and protein synthesis (Ghassemi-Golezani and Taifeh-Noori 2011). Salinity-related water stress has negative effect on the nodulation ability of soybean plants. It is now recognized that different types of signals (nitrogenous, oxidative, redox, etc.) in legume's rhizospheric environment play a crucial role in establishing interaction

between microbes and legumes (Karmakar et al. 2015). Therefore, soybean is a useful model for investigating the ecology of PGPR and their influence on plant growth under saline soils. This study was conducted to assess the PGPR that are associated with the rhizosphere of soybean grown in semiarid areas of Rajasthan. We also sought to identify and characterize representative PGPR with respect to growth-promoting attributes and studied their salinity tolerance. Rhizosphere environment is hot spot for various types of soil microorganisms due to rich nutrient availability. The bacteria present in stress affected rhizospheric environments have been isolated and identified and their beneficial effects on plant growth studied by many researchers (Jain et al. 2014; Vaishnav et al. 2015; Kumari et al. 2015).

Lane 1- DNA marker (100 bp), Lane 2- *g6pd*, Lane 3- *p5cr*, Lane 4- *nr,* Lane 5- *amy*, Lane 6- *IaaM*, Lane 7- *sid,* Lane 8- *acds*

The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. Values shown next to the branches are the percentage of replicate trees with associated taxa clustered together in the bootstrap test (1000 replicates). Evolutionary analyses were conducted in MEGA 6.

The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. Values shown next to the branches are the percentage of replicate trees with associated taxa clustered together in the bootstrap test (1000 replicates). Evolutionary analyses were conducted in MEGA 6.

The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. Values shown next to the branches are the percentage of replicate trees with associated taxa clustered together in the bootstrap test (1000 replicates). Evolutionary analyses were conducted in MEGA 6.

In all tested bacteria, the growth was reduced upon higher NaCl and PEG concentration. More than 65% of bacterial strains were severely affected by 8% NaCl and 10% PEG. Only a few isolates such as AT4, AT12, AM3, AM5 and AU showed a significant growth with 10% NaCl and 12% PEG. These results are in agreement with several studies, which showed that the number of viable colonies of rhizospheric strains declined with increasing concentration of salt, while isolated strains from saline soil were able to survive at higher salt concentration in media (Naz et al. 2009; Omar et al. 2009). A total number of 14 bacterial isolates which were showed significant growth on higher salt stress were screened in vitro for PGP activities. A number of three bacterial isolates were found to show all PGP properties tested, whereas six bacterial isolates were able to produce more than one PGP properties.

A total number of eight bacterial isolates were able to produce IAA in the presence of tryptophan; however, out of them four isolates were also capable to produce IAA in the absence of tryptophan. Various reports have been published on IAA production by PGPR strains either with or without the tryptophan supplement in culture media (Farina et al. 2012; Mia and Shamsuddin 2013). In our experiment the level of IAA produced in culture media was found higher with tryptophan. Strain

AU was found to have higher ability to convert tryptophan to IAA than other isolates, while in the absence of tryptophan, its ability was similar to AM3 isolate. The production of IAA by PGPR is usually correlated with direct effects on plant growth (Cassán et al. 2014). It is responsible for division, extension and differentiation of plant cells and tissues. It also affects photosynthesis, pigment formation, biosynthesis of various metabolites and resistance to biotic and abiotic stress factors (Bashan and de-Bashan 2010). It has been proved that IAA participates in biofilm formation leading to quorum sensing (Hu et al. 2010). Inoculation of IAA producing *Carnobacterium* sp. was reported to increase soybean seedling growth against *Fusarium* wilt (Jain et al. 2013). The IAA producing wild and mutant type of *Pseudomonas* and *Bacillus* strains were found to enhance tolerance against salinity and drought stress in soybean and mung bean plants, respectively (Kumari et al. 2016). Only seven isolates were able to solubilize $Ca_3(PO_4)_2$ present in the Pikovskaya medium in salt stress as well as non-salt stress. The capacity to solubilize Pi is a good characteristic for the selection of bacteria capable of increasing P content in the rhizosphere, and their application to increase plant protection against adverse abiotic factors has now been an upcoming strategy (reviewed by Vassilev et al. 2012). The solubilization of calcium phosphate observed in this study may be due to the production of organic acids by the bacteria which decreases pH in the culture medium (Marra et al. 2012). Similarly phosphate-solubilizing bacteria have been isolated from sunflower (Ambrosini et al. 2012), canola (Farina et al. 2012) and Turkish tea (Çakmakçı et al. 2010). Phosphate-solubilizing bacteria (PSB) isolated from canola plants were tested for growth-promoting effects and found to promote plant growth (Farina et al. 2012). Likewise novel PSB, '*Pantoea cypripedii* PS1' along with *Enterobacter aerogenes* PS16 and *Rhizobium ciceri* enhanced the growth of chickpea (*Cicer arietinum*L.) (Singh et al. 2014). Besides, *Burkholderia* was found as a dominant rhizospheric bacterial genus associated with sunflower plants and showed a stimulatory effect on plant growth through siderophore and solubilized phosphate (Ambrosini et al. 2012).

Another important trait of rhizospheric bacteria is production of siderophores that may directly or indirectly influence the plant growth. Directly, plant roots uptake iron from siderophore, and indirectly, siderophores are making iron unavailable to the phytopathogens and protecting plant health (Rajkumar et al. 2010). In the present study, five isolates were found positive on CAS medium and produced siderophore up to 8% NaCl. Likewise, a total of 108 bacteria isolated from canola plants displayed the ability to produce siderophores in which two bacterial genus *Klebsiella* and *Pseudomonas* were able to promote canola plant growth (Farina et al. 2012). Siderophore-producing *Paenibacillus* sp. isolates were recovered from wheat rhizosphere and chosen for in vivo experiments in a greenhouse and found to be very efficient in wheat plant growth promotion (Beneduzi et al. 2008).

The ACC-D-containing PGPRs have been reported for improving the salinity tolerance in plants by many researchers (Siddikee et al. 2011; Nautiyal et al. 2013; Kumari et al. 2015; Singh et al. 2015). A number of six bacterial isolates were

exhibited ACC-D activity in which only three isolates AU, AU7 and AT11 were able to show activity up to 8% NaCl. The inoculation of *Klebsiella* sp. SBP-8 containing ACC-deaminase on wheat (*Triticum aestivum*) under salt condition improved growth of the plant and protects from salt stressors through more than one mechanism including an effect on plant biomass, chlorophyll content and on K^+/Na^+ ratio (Singh et al. 2015). ACC deaminase containing *Pseudomonas* sp. mediated saline tolerance in soybean (*Glycine max*) plants (Kumari et al. 2015) and drought stress in drying soil in mung bean (*Vigna radiata*) (Kumari et al. 2016).

Introduction of EPS-producing microorganisms in the drought and saline environment can alleviate stress in the crop plants. In this study, salt-tolerant EPS-producing rhizobacteria from soybean rhizosphere were screened. Only seven bacterial isolates produced mucoid growth. These isolates showed efficient EPS production up to 8% NaCl. The EPS production was higher in the presence of NaCl stress than non-stress condition, and it increased by increasing stress level in the most of the isolates, indicating that EPS production in bacteria occurs as a response to the stress and protects cells against desiccation (Qurashi and Sabri 2011). EPS-producing bacteria in the root zone were reported to enhance soil aggregation and water holding capacity and decrease the content of Na^+ available for plant uptake and thus help in alleviating salt and drought stress in plants (Choudhary et al. 2015). An EPS-producing bacterial strain *Pseudomonas putida* GAP-P45 was recovered from alfisol of sunflower rhizosphere and found to increase the survival, plant biomass and root adhering soil/root tissue ratio of sunflower seedlings under drought stress (Vardharajula et al. 2009). Upadhyay et al. (2011) isolated 11 bacterial strains from wheat grown on salt-affected soils, which showed tolerance up to 80 g L^{-1}NaCl and also exhibited an EPS-producing potential. The isolated bacterial strains were employed as inocula in wheat plants and found to increase biomass compared to the uninoculated plants during salt stress.

Pseudomonads are predominant bacteria in the rhizosphere with versatile functions. They are known to produce plant hormones, siderophores, antibiotics, enzymes like proteases and glucanases and solubilize minerals which have made them the most promising group of PGPR involved in the biocontrol of plant diseases and abiotic stress tolerance (reviewed by Choudhary 2012). They were capable of rapid growth and utilize various substrates as nutrients therefore show efficient colonization with a wide variety of crops including cereals, pulses, oilseeds and vegetables (Santoyo et al. 2012). In addition, PCR amplification confirmed the presence of PGP genes, namely, *acdS, nr, amy, IaaM, sid, g6pd* and *p5cs*in bacterial isolate AU. Presence of ACC-D has been confirmed by amplification and sequence analysis of *acdS*, a structural gene encoding ACC-D. Several pair of primers has been designed by various researchers to detect the presence of *acdS* gene in bacteria (Hontzeas et al. 2004; Cheng et al. 2008; Duan et al. 2009; Jha et al. 2012). The *acdS* gene is commonly found in *Actinobacteria*, *Deinococcus-Thermus*, three classes of *Proteobacteria* (α, β and γ), various fungi belonging to *Ascomycota* and *Basidiomycota* and in some Stramenopiles (Singh et al. 2015).

In the study performed by our group, an inoculation of PGPB *Pseudomonas* spp. to soybean and guar in the 200 mMNaCl treatment alleviated the stress (Kumari et al. 2016, 2016; Vaishnav et al. 2015). All the treatment plants with bacterial inoculation showed better plant growth characteristic with respect to their control. Salt-treated plants showed least growth due to salinity-induced harmful physiological changes in plant system, while the plants inoculated with bacterial strains showed good growth characteristics. On comparing the results of *Pseudomonas* spp. inoculated soybean and guar plants, it was found that the strain AU was able to promote the growth of stress-tolerant guar plants along with soybean plants. Being stress tolerant guar showed more percentage seed germination than soybean plants, but the effect of PGPB can be clearly seen on the plants. Plant growth-promoting activities of bacteria make them efficient for alleviating biotic as well as abiotic stress (Choudhary et al. 2015).

Acknowledgments The research was supported by SERB-Grant no. SR/FT/LS-129/2012 to DKC. Some of the research has partially been supported by DBT grant no. BT/PR1231/AGR/21/340/2011 to DKC.

References

Algar E, Gutierrez-Mañero FJ, Garcia-Villaraco A, García-Seco D, Lucas JA, Ramos-Solano B (2014) The role of isoflavone metabolism in plant protection depends on the rhizobacterial MAMP that triggers systemic resistance against *Xanthomonas axonopodis* pv. *glycines* in *Glycine max* (L.) Merr. cv. Osumi. Plant Physiol Biochem 82:9–16

Ambrosini A, Beneduzi A, Stefanski T, Pinheiro FG, Vargas LK, Passaglia LM (2012) Screening of plant growth promoting Rhizobacteria isolated from sunflower (*Helianthus annuus* L.). Plant Soil 356(1–2):245–264

Armendariz AL, Talano MA, Wevar Oller AL, Medina MI, Agostini E (2015) Effect of arsenic on tolerance mechanisms of two plant growthpromoting bacteria used as biological inoculants. J Environ Sci 33:203–210

Bashan Y, De-Bashan LE (2010) How the plant growth-promoting bacterium *Azospirillum* promotes plant growth: a critical assessment. Adv Agron 108:77–136

Beneduzi A, Peres D, da Costa PB, Zanettini MHB, Passaglia LMP (2008) Genetic and phenotypic diversity of plant-growth-promoting bacilli isolated from wheat fields in southern Brazil. Res Microbiol 159:244–250

Bhattacharyya PN, Jha DK (2012) Plant growth-promoting rhizobacteria (PGPR): emergence in agriculture. World J Microbiol Biotechnol 28(4):1327–1350

Çakmakçı R, Dönmez MF, Ertürk Y, Erat M, Haznedar A, Sekban R (2010) Diversity and metabolic potential of culturable bacteria from the rhizosphere of Turkish tea grown in acidic soils. Plant Soil 332:299–318

Cassán F, Vanderleyden J, Spaepen S (2014) Physiological and agronomical aspects of phytohormone production by model plant-growth-promoting rhizobacteria (PGPR) belonging to the genus *Azospirillum*. J Plant Growth Regul 33:440–459

Cheng Z, Duncker BP, McConkey BJ, Glick BR (2008) Transcriptional regulation of ACC deaminase gene expression in *Pseudomonas putida* UW4. Can J Microbiol 54:128–136

Choudhary DK (2012) Microbial rescue to plant under habitat-imposed abiotic and biotic stresses. Appl Microbiol Biotechnol 96(5):1137–1155

Choudhary DK, Kasotia A, Jain S, Vaishnav A, Kumari S, Sharma KP, Varma A (2015) Bacterial-mediated tolerance and resistance to plants under abiotic and biotic stresses. J Plant Growth Regul 1:25

Choudhary DK, Kasotia A, Jain S, Vaishnav A, Kumari S, Sharma KP, Varma A (2016) Bacterial-mediated tolerance and resistance to plants under abiotic and biotic stresses. J Plant Growth Regul 35:276–300

Compant S, Duffy B, Jerzy N, Clement C, Barka EA (2005) Use of plant growth-promoting bacteria for biocontrol of plant diseases: principles, mechanisms of action, and future prospects. Appl Environ Microbiol 71(9):4951–4959

Duan J, Müller KM, Charles TC, Vesely S, Glick BR (2009) 1- Aminocyclopropane-1- carboxylate (ACC) deaminase genes in *Rhizobia* from southern Saskatchewan. Microb Ecol 57:423–436

Egamberdieva D, Shurigin V, Gopalakrishnan S, Sharma R (2015) Microbial strategies for the improvement of legume production in hostile environments. In: Azooz MM, Ahmad P (eds) Legumes under environmental stress: yield, improvement and adaptations, 1st edn. Wiley, Hoboken

FAO (2013) Climate-smart agriculture sourcebook

Farina R, Beneduzi A, Ambrosini A, de Campos SB, Lisboa SS, Wendisch V, Vargas LK, Passaglia LMP (2012) Diversity of plant growth-promoting rhizobacteria communities associated with the stages of canola growth. App Soil Ecol 55:44–52

Fernández-Bidondo L, Silvani V, Colombo R, Pérgola M, Bompadre J, Godeas A (2011) Pre-symbiotic and symbiotic interactions between Glomus intraradices and two *Paenobacillus* species isolated from AM propagules. In vitro and in vivo assays with soybean (AG043RG) as plant host. Soil Biol Biochem 43:1866–1872

Ghassemi-Golezani K, Taifeh-Noori M (2011) Soybean performance under salinity stress. INTECH Open Access Publisher, Rijeka

Holt JG, Krieg NR, Sneath PH, Staley JT, Williams ST (1994) International edition: Bergey's manual of determinative bacteriology. Springer Science+Business Media, New York

Hontzeas N, Zoidakis J, Glick BR, Abu-Omar MM (2004) Expression and characterization of the 1-aminocyclopropane-1-carboxylic acid deaminase from the rhizobacterium *P. putida* UW4: a key enzyme in bacterial plant growth promotion. Biochem Biophys Acta 1703:11–19

Hu M, Zhang C, Mu Y, Shen Q, Feng Y (2010) Indole affects biofilm formation in bacteria. Ind J Microbiol 50(4):362–368

Jain S, Vaishnav A, Kasotia A, Kumari S, Gaur RK, Choudhary DK (2013) Bacteria-induced systemic resistance and growth promotion in Glycine max L. Merrill upon challenge inoculation with Fusarium oxysporum. Proc Natl Acad Sci India Biol Sci 83:561–567

Jain S, Vaishnav A, Kasotia A, Kumari S, Gaur RK, Choudhary DK (2014) Bacteria-induced systemic resistance and growth promotion in *Glycine max* L. Merrill upon challenge inoculation with *Fusarium oxysporum*. Proc Natl Acad Sci India Biol Sci 83:561–567

Jha B, Gontia I, Hartmann A (2012) The roots of the halophyte *Salicornia brachiata* are a source of new halotolerant diazotrophic bacteria with plant growth-promoting potential. Plant Soil 356:265–277

Juge C, Prévost D, Bertrand A, Bipfubusa M, Chalifour FP (2012) Growth and biochemical responses of soybean to double and triple microbial associations with *Bradyrhizobium*, *Azospirillum* and arbuscular mycorrhizae. Appl Soil Ecol 61:147–157

Karmakar K, Rana A, Rajwar A, Sahgal M, Johri BN (2015) Legume-rhizobia symbiosis under stress. In: Arora NK (ed) Plant microbes Symbiosis: applied facets, vol 241. Springer, India. https://doi.org/10.1007/978-81-322-2068-8_12

Khodair TA, Galal GF, El-Tayeb TS (2008) Effect of inoculating wheat seedlings with exopolysaccharide-producing bacteria in saline soil. J Appl Sci Res 4:2065–2070

Kumari S, Vaishnav A, Jain S, Varma A, Choudhary DK (2015) Bacterial-mediated induction of systemic tolerance to salinity with expression of stress alleviating enzymes in soybean (*Glycine max* L. Merrill). J Plant Growth Regul 34:558–573

Kumari S, Vaishnav A, Jain S, Varma A, Choudhary DK (2016) Induced drought tolerance through wild and mutant bacterial strain *Pseudomonas simiae* in mung bean (*Vigna radiata* L.). World J Microbiol Biotechnol 32(1): 1–0

Marks BB, Megías M, Nogueira MA, Hungria M (2013) Biotechnological potential of rhizobial metabolites to enhance the performance of *Bradyrhizobium* spp. and *Azospirillum brasilense* inoculants with soybean and maize. AMB Express 3:21

Marra LM, Soares CRFS, de Oliveira SM, Ferreira PAA, Soares BL, de Fráguas CR, de Lima JM, de Souza Moreira FM (2012) Biological nitrogen fixation and phosphate solubilization by bacteria isolated from tropical soils. Plant Soil 357:289–307

Masciarelli O, Llanes A, Luna V (2014) A new PGPR co-inoculated with *Bradyrhizobium japonicum* enhances soybean nodulation. Microbiol Res 169:609–615

Mia MB, Shamsuddin ZH (2013) *Rhizobium* as a crop enhancer and biofertilizer for increased cereal production. Afr J Biotechnol 9:6001–6009

Munns R, Tester M (2008) Mechanisms of salinity tolerance. Annu Rev Plant Biol 59:651–681

Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K (2014) The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. Front Plant Sci. https://doi.org/10.3389/fpls.2014.00170

Nautiyal CS, Srivastava S, Chauhan PS, Seem K, Mishra A, Sopory SK (2013) Plant growth-promoting bacteria *Bacillus amyloliquefaciens* NBRISN13 modulates gene expression profile of leaf and rhizosphere community in rice during salt stress. Plant Physiol Biochem 66:1–9

Naz I, Bano A, Hassan T-U (2009) Isolation of phytohormones producing plant growth promoting rhizobacteria from weeds growing in Khewra salt range, Pakistan and their implication in providing salt tolerance to *Glycine max* L. Afr J Biotechnol 8(21):5762–5766

Omar MNA, Osman MEH, Kasim WA, El-Daim A (2009) Improvement of salt tolerance mechanisms of barely cultivated under salt stress using Azospirillum brasilense. In: Ashraf M, Ozturk M, Athar HR (eds) Salinity and water stress, vol 44. Springer, Netherlands, pp 133–147

Prudent M, Vernoud V, Girodet S, Salon C (2015) How nitrogen fixation is modulated in response to different water availability levels and during recovery: a structural and functional study at the whole plant level. Plant Soil 399:1–12

Qurashi AW, Sabri AN (2011) Osmoadaptation and plant growth promotion by salt tolerant bacteria under salt stress. Afr J Microbiol Res 5:3546–3554

Rajkumar M, Ae N, Prasad MN, Freitas H (2010) Potential of siderophore-producing bacteria for improving heavy metal phytoextraction. Trends Biotechnol 28(3):142–149

Ramesh A, Sharma SK, Yadav N, Joshi OP (2014a) Phosphorus mobilization from native soil p-pool upon inoculation with phytate-mineralizing and phosphate solubilizing *Bacillus aryabhattai* isolates for improved p-acquisition and growth of soybean and wheat crops in microcosm conditions. Agric Res 3:118–127

Ramesh A, Sharma SK, Sharma MP, Yadav N, Joshi OP (2014b) Inoculation of zinc solubilizing *Bacillus aryabhattai* strains for improved growth, mobilization and biofortification of zinc in soybean and wheat cultivated in Vertisols of Central India. Appl Soil Ecol 73:87–96

Salavati A, Bushehri AAS, Taleei A, Hiraga S, Komatsu S (2012) A comparative proteomic analysis of the early response to compatible symbiotic bacteria in the roots of a supernodulating soybean variety. J Proteome 75:819–832

Santoyo G, Orozco-Mosqueda MDC, Govindappa M (2012) Mechanisms of biocontrol and plant growth-promoting activity in soil bacterial species of *Bacillus* and *Pseudomonas*: a review. Biocontrol Sci Technol 22:855–872

Schubert S, Neubert A, Schierholt A, Su¨mer A, Zörb C (2009) Development of salt-resistant maize hybrids: the combination of physiological strategies using conventional breeding methods. Plant Sci 177:196–202

Shafique A, Rehman S, Khan A, Kazi AG (2014) Improvement of legume crop production under environmental stresses through biotechnological intervention. In: Ahmad P (ed) Emerging

Technologies and Management of Crop Stress Tolerance, vol 2. https://doi.org/10.1016/B978-0-12-800875-1.00001-6

Shrivastava P, Kumar R (2015) Soil salinity: a serious environmental issue and plant growth promoting bacteria as one of the tools for its alleviation. Saudi J Biol Sci 22:123–131

Siddikee MA, Glick BR, Chauhan PS, Yim W, Sa T (2011) Enhancement of growth and salt tolerance of red pepper seedlings (*Capsicum annuum* L.) by regulating stress ethylene synthesis with halotolerant bacteria containing 1-aminocyclopropane-1-carboxylic acid deaminase activity. Plant Physiol Biochem 49:427–434

Simonetti E, Viso NP, Montecchia M, Zilli C, Balestrasse K, Carmona M (2015) Evaluation of native bacteria and manganese phosphite for alternative control of charcoal root rot of soybean. Microbiol Res 180:40–48

Singh O, Gupta M, Mittal V, Kiran S, Nayyar H, Gulati A, Tewari R (2014) Novel phosphate solubilizing bacteria '*Pantoea cypripedii* PS1'along with *Enterobacter aerogenes* PS16 and *Rhizobium ciceri* enhance the growth of chickpea (*Cicer arietinum* L.). Plant Growth Regul 73:79–89

Singh RP, Jha P, Jha PN (2015) The plant-growth-promoting bacterium *Klebsiella* sp. SBP-8 confers induced systemic tolerance in wheat (*Triticum aestivum*) under salt stress. J Plant Physiol 184:57–67

Stefan M, Dunca S, Olteanu Z, Oprica L, Ungureanu E, Hritcu L, Mihasan M, Cojocaru D (2011) Soybean (*Glycine max* [L] Merr.) inoculation with *Bacillus pumilus* RS3 promotes plant growth and increases seed protein yield: relevance for environmentally-friendly agricultural applications. Carpathian J Earth Environ Sci 5:131–138

Sun L, Qiu FB, Zhang XX, Dai X, Dong XZ, Song W (2008) Endophytic bacterial diversity in rice (*Oryza sativa* L.) roots estimated by 16S rDNA sequence analysis. Microb Ecol 55:415–424

Upadhyay SK, Singh JS, Singh DP (2011) Exopolysaccharide-producing plant growth-promoting rhizobacteria under salinity condition. Pedosphere 21(2):214–222

USDA (2015) World agriculture production

Vaishnav A, Kumari S, Jain S, Varma A, Choudhary DK (2015) Putative bacterial volatile-mediated growth in soybean (*Glycine max* L. Merrill) and expression of induced proteins under salt stress. J Appl Microbiol 119:539–551

Vardharajula S, Ali SKZ, Grover M, Reddy G, Bandi V (2009) Alleviation of drought stress effects in sunflower seedlings by the exopolysaccharides producing *Pseudomonas putida* strain GAP-P45. Biol Fertil Soils 46:17–26

Vassilev N, Eichler-Löbermann B, Vassileva M (2012) Stress-tolerant P-solubilizing microorganisms. Appl Microbiol Biotechnol 95:851–859

Wang C-J, Yang W, Wang C, Gu C, Niu D-D, Liu H-X, Wang Y-P, Guo J-H (2012) Induction of drought tolerance in cucumber plants by a consortium of three plant growth-promoting rhizobacterium strains. PLoS One 12:1–10

Chapter 3
Methanogenic Archaea in Paddy Agricultural Fields

Navnita Srivastva, Alpana Singh, and Suresh K. Dubey

Abstract The analysis of methane-producing methanogenic archaea community structure will be helpful in enhancing the understanding of microbial ecology of methanogens and methane production potential of soil that is essential for formulation of strategies to mitigate methane emission from rice fields. Hence, the present study aimed to explore the methanogenic archaea community structure and their methane production potential in three different rice soil types (Alfisol, Inceptisol, and Vertisol). Phylogenetic analysis revealed the presence of *Methanocellales*, *Methanomicrobiaceae*, *Methanobacteriaceae*, *Methanosarcinaceae*, and *Methanosaetaceae* group common to all the soils with varying relative abundance. Methane production potential for soils was in order Alfisol < Inceptisol < Vertisol, following the first-order kinetics. Overall analyses suggested that there is homogeneity in soils for methanogenic diversity, although methane production potential varied, possibly due to the presence of methanogens in different proportions and difference in the soil characteristics.

Keywords Paddy soil · Methanogens · Methane production · Kinetics · Diversity

3.1 Introduction

Methane is the main hydrocarbon present in the atmosphere, with an average concentration of 1803 ppb (IPCC 2013). Its ability to absorb infrared radiation makes it a potent greenhouse gas, with a warming potential 20–30 times more than that of CO_2 (Blake and Rowland 1988; Rodhe 1990). Each year, 50 million tonnes of methane is added to the atmosphere, resulting in a relative annual increase of more than 1% (Bouwman 1990). It is produced from many natural (wetlands, termites, oceans, fresh water, mud volcanoes, livestock, etc.) and anthropogenic (fossil fuels, rice paddies, biomass burning, landfills, animal waste, domestic

N. Srivastva · A. Singh · S. K. Dubey (✉)
Department of Botany, Institute of Science, Banaras Hindu University, Varanasi, India
e-mail: skdubey@bhu.ac.in

D. K. Choudhary et al. (eds.), *In Silico Approach for Sustainable Agriculture*,
https://doi.org/10.1007/978-981-13-0347-0_3

sewage, etc.) sources (Uprety et al. 2011). Among all anthropogenic sources, rice ecosystem is one of the most important source of methane emission and accounts for the global methane budget of about 21% (Prather and Ehhalt 2001; Chen and Prinn 2006). The contribution from rice cultivar is estimated to range from 20 to 100 Tg CH_4 $year^{-1}$ with an average of 60 Tg CH_4 $year^{-1}$ (IPCC 1996). According to IPCC estimates (IPCC 1995), production of 1 kg of rice corresponds to the emission of 100 g of methane. Methane emission from rice fields is exposed to intensify in the future due to increase of rice production in order to supply the food demand for rapidly growing population of world. It is estimated that the methane emission from rice cultivation would rise to 145 Tg $year^{-1}$ by 2025 (Anastasi et al. 1992).

Emission of methane from irrigated rice fields is the result of complex interactions between rice plants and the microbial community of the soil (Neue and Sass 1994; Conrad 1996). Root exudation supports the methanogenic food chain in the rhizosphere and is thus responsible for enhanced methane production (Aulakh et al. 2001). Methane production potential of soil also depends on the quality and the amount of C-substrate available to microbial community (Hou et al. 2000). Plant roots release about 17% of the photosynthate captured, most of which is available to soil organisms (Nguyen 2003). Microbial populations in rice soil differ greatly in their responses to plant photosynthate input (Lu et al. 2004). Methanogens lack enzymes responsible for the breakdown of such organic matter and thus depend on other syntrophs to decompose complex organic biopolymers to simple substrates (Zinder 1993). In the initial process of decomposition, the fermenting bacteria secrete enzymes for the hydrolysis of polysaccharides and result in the formation of sugar monomers to alcohols, fatty acids, and H_2 (Chin et al. 1999). Further, secondary fermentation is carried out by syntrophic microorganisms to produce acetate, CO_2, and H_2. Other physiological group of fermenting bacteria, i.e., homoacetogenic bacteria, ferments sugars directly into acetate (Drake 1994). Some of the homoacetogens are able to convert H_2 + CO_2 to acetate. Methanogens utilize some of the fermentation products, i.e., acetate, hydrogen, and other methyl compounds on degradation, and produce CH_4 at the terminal step of decomposition.

Thus, the methane emissions by soil result from coordinated and correlated microbial activities (le Mer and Roger 2001). Methane is generated in the anaerobic zones of submerged rice field soils by methanogens belonging to phylum *Euryarchaeota* of *Archaea* (Woese et al. 1990). On the basis of 16S rRNA oligonucleotide sequences, membrane lipid composition, and antigenic fingerprinting data, the methanogens have been taxonomically classified into seven orders, *Methanobacteriales*, *Methanococcales*, *Methanomicrobiales*, *Methanosarcinales*, *Methanopyrales*, *Methanocellales*, and *Methanomassiliicoccales* (Borrel et al. 2014). They are important catalysts in the global carbon cycle. They are responsible for processing 1–2% of the carbon fixed every year during photosynthesis (Reeburgh et al. 1993). They are strictly anaerobic archaea which produces methane as one of the final products of energy metabolism. They convert a few simple compounds such as H_2 + CO_2, formate, methanol, methylamines and acetate to methane. Methanogenesis from all these substrates requires unique coenzymes, some of which are exclusively found in methanogens (Ludmila et al. 1998). About 77% of

methanogenic species are hydrogenotrophic, in which 60% also utilize formate, 14% are acetotrophic, and 28% are methylotrophic (Garcia 1990; le Mer and Roger 2001).

The distribution of methanogens in natural environment is highly dependent on their adaptation to various temperature, pH, and salinity ranges (Garcia 1990). Carbon availability, temperature, moisture, pH, redox potential (E_h), nutrient dynamics, and inorganic redox ions are the key variables controlling methane emission from anaerobic soil systems (Wang et al. 1993; Sass et al. 1994; Wagner et al. 1999; Kumaraswamy et al. 2001). Information on the effects of different soil parameters on methane emission and methanogens is necessary to provide a theoretical basis for controlling methane emission in rice soil.

Several studies, using different culture-independent molecular techniques, such as T-RFLP (terminal restriction fragment length polymorphism) (Kruger et al. 2005; Kniffin et al. 2010), cloning (Ramakrishnan et al. 2001), DGGE (denaturing gradient gel electrophoresis), and DNA stable-isotope probing (Watanabe et al. 2006, 2011), have been conducted to analyze diversity of methanogens in paddy soils. These techniques target the conserved gene sequences such as small subunit rRNA (16S rRNA) or the methanogens-specific *mcr*A gene, for identification of methanogens present in different habitats. Variations have been observed in methanogenic community structure in paddy fields of different countries. Wang et al. (2010) explore different groups of methanogenic community from the paddy soil located in northeast China with different characteristics and found variation due to soil type and sampling location. The information on community structure of methanogens in paddy field soils is available for Japan (Watanabe et al. 2006), Italy (Kruger et al. 2005) and Philippines (Ramakrishnan et al. 2001). Hoshino et al. (2011) also reported the effect of soil type on archaeal community structure (diversity and density) in Japanese paddy soil. Cao et al. (2012) reported a differential distribution of archaeal community for different soils and suggested that the difference in the copy number is positively correlated with pH, organic C, and total N. Conrad et al. (2012) have determined the methanogenic population dynamics for the degradation process of straw in anoxic condition of three types of soils belonging to different countries, *viz.*, Italy, China, and Thailand. These studies, thus, signify the effect of geographical conditions on methanogen community structure. However, most of these investigations are from rice fields which are confined to temperate ecozones. The information of methanogenic diversity in tropical soil is inadequate.

Methane production potential of soil by several field measurement and laboratory incubation studies has also been reported (Kruger et al. 2005; Conrad et al. 2006). Few studies on methane flux (Jia et al. 2006) for tropical soil are available. To know the effect of soil properties on methane emission, 16 different rice fields were selected from the USA, Thailand, India, and Liberia. They found that only soil clay content is positively correlated to CH_4 mitigation (Wang et al. 1992). However, the data on CH_4 fluxes are available from different types of soils, but the way by which soil characteristics control the methane emission is still not clear.

Among the rice-growing countries, India has the largest cultivated area with a variety of soil types (about 45 million ha.) and ranks next to China (USDA FAS

2009). Yan et al. (2003) reported that out of the total 28.2 Tg CH_4 emission per year from rice fields worldwide, Asian countries contribute 25.1 Tg CH_4 with 5.88 Tg CH_4 from India alone. In India, four higher methane-emitting or "hot spot" states (West Bengal, Bihar, Madhya Pradesh, and Uttar Pradesh) have accounted for 53.9% of total CH_4 emission (Gupta et al. 2009). In light of the fact that soil characteristics regulate the microbial community structure, which in turn affects methanogenesis, studies on methanogenic community seem imperative in understanding the ecology of methanogens and the control of methane turnover in rice soils. Methane flux data available for Indian rice fields (Singh et al. 1999; Kumaraswamy et al. 2001; Rath et al. 2005) also supports the correlation between soil physicochemical properties and CH_4 emission. Earlier investigations on Indian rice fields (Singh et al. 1999; Adhya et al. 2000; Anand et al. 2005) have revealed that tropical agroecosystems are the main source of CH_4. The effect of irrigation and manure practices was studied for Indian paddy fields by Pathak et al. (2003) in order to apply the mitigation options for methane emissions. But these studies are not focused on microbiological aspect of CH_4 production in rice fields. Apart from these studies, Conrad (2007) also reported India as the total unexplored country in context of methanogenic community dwelling in paddy fields. Thus, the information on diversity of methanogenic archaea in Indian paddy soil is almost negligible. Hence, in the present study, we investigated the effects of different soil types on methanogen community structure and methane production potential.

3.2 Materials and Methods

3.2.1 Experimental Site and Soil Sampling

Soil samples of cropped bulk (between the plant rows) were collected in triplicate from N-fertilized (urea applied at the rate of 30 kg N ha^{-1}) farmers' rice fields at three different locations of eastern Uttar Pradesh, India, in the month of September, 2011. Three different sites were located approximately 60 km apart from each other. Cultivation practices and fertilizer application were common. The soil samples were collected at five different points and pooled together to make a composite sample to avoid any heterogeneity and to represent the overall characteristics of the sites as reported earlier (Vishwakarma et al. 2010). Composite soil samples were further divided into two subsamples, one of fresh samples stored at −20 °C for molecular analyses and the other air dried for soil physicochemical characteristics. All analyses were in triplicate. Soil samples were collected from 0 to 10 cm depth using a 5 cm diameter soil corer during a week and sieved through 2 mm mesh. Fields were irrigated, and N-fertilizer (urea) was also applied at the rate of 30 kg N ha^{-1} before 1 week of soil sampling as per farmers' practice.

Table 3.1 Sampling site description and soil characteristics

Parameters	Soil types		
	Alfisol	Inceptisol	Vertisol
Longitude	83°34′20.479″ E	83°59′8.964″ E	83°15′49.04″ E
Latitude	25°9′9.574″ N	25°16′31.468″ N	25°14′46.81″ N
Soil type	Alfisol	Inceptisol	Vertisol
Texture (%)	Sand 34, silt 64, clay 2	Sand 32, silt 65, clay 3	Sand 6, silt 82, clay 12
pH	6.2	7.3	7.5
E_h (mV)	−60	−67.5	−92.2
WHC (%)	34.1	43.5	45.0
Total N (%)	0.12 ± 0.01	0.12 ± 0.03	0.16 ± 0.13
Organic C (%)	0.57 ± 0.06	0.75 ± 0.12	0.85 ± 0.04

"±" indicate the standard error

3.2.2 Soil Physicochemical Analyses

Soil texture, WHC, organic C, total N, and NH_4^+-N were determined as per the standard procedure (Walkey 1947; Jackson 1958; APHA 1985; Singh et al. 2012). Redox potential (E_h) of soils was measured using digital potentiometer equipped with Pt-electrode (Systronics, India) (Yao et al. 1999). The soil characteristics of the study sites are given in Table 3.1.

3.2.3 Methane Production Potential of Soil

To study methane production, 100 mL of anoxic sterile water was added to 50 g of soil in a 250 mL flask. The flask was closed and sealed with a butyl rubber stopper. The stopper was specially designed for the experiment, having two openings to allow N_2 flushing and a third opening with rubber septum to allow sampling of gas. Before incubation the headspace was purged with O_2-free N_2 gas for at least 3 min with constant shaking of flask to promote anaerobic environment. The flasks were incubated statically in dark at 30 °C for 20 days (Jia et al. 2006). All experiments were carried out in triplicate. Methane concentration in the headspace was measured with a gas chromatograph by sampling 1 mL of accumulated gas in the headspace after vigorous shaking of flask by hand to allow equilibration between the liquid and gas phases. The methane concentration was analyzed by the gas chromatograph (Varian 3800, The Netherland), equipped with flame ionization detector (FID), using 3 m Porapak Q column. The oven, injector, and detector temperatures were set as 80 °C, 70 °C, and 200 °C, respectively. The flow rate of carrier gas N_2 and flame gases H_2 and O_2 were set as 30 ml min^{-1}, 20 ml min^{-1}, and 300 ml min^{-1}, respectively. Methane production potential of each soil sample was calculated from the CH_4 concentration in the headspace measured at intervals of 0, 60, 120, 180, 240, 300, 360, 420, and 480 h.

3.2.4 Kinetics of Methane Production

The methane volume at different time intervals for various soil samples was used to evaluate the kinetic constants for methane production using the first-order rate expression obtained from the Monod model.

$$\frac{d[M]}{dt} = K' \frac{\mu_{\max}[X][S]}{K_s + [s]} \tag{3.1}$$

For $K_s >> [S]$, and constant [X] as

$$K = K' \frac{\mu_{\max} X_{const}}{K_s} \tag{3.2}$$

Hence,

$$\frac{d[M]}{dt} = K[S] \tag{3.3}$$

where $K' = 1/Y$ is the increase of yield coefficient, μ_{max} is the maximum specific growth rate, $[X]_{const}$ is the methanogenic population number, and K_s is the substrate constant. Since the concentration and nature of substrate are not fully known, the substrate concentration [S] can be replaced by $[M_m - M]$ to get the following equation:

$$\frac{d[M]}{dt} = K[M_m - M] \tag{3.4}$$

where constant K is a pseudo-constant obtained from the Monod model.

3.2.5 PCR-Cloning-RFLP Analysis, Sequencing, and Assignment of Clones into OTUs

Total genomic DNA from soil samples (0.5 g) was extracted using FastDNA® Spin Kit for Soil (MP Biomedicals, LLC. Solon, Ohio) according to the manufacturer's protocol. The methanogen-specific 16S rRNA genes were amplified by PCR with the primer sets 0357 F (5′-CCCTACGGGGCGCAGCAG-3′) and 0691 R (5-′-GGATTACARGATTTCAC-3′) (Watanabe et al. 2004). The reaction mixture contained 2.5 μL of 10× buffer (Fermentas, UK), 100 μM of each dNTPs (Fermentas, UK), 0.5 μM of each primer (Sigma, USA), and 1 U of *Taq DNA polymerase* (New England Biolabs, Beverly, MA). Amplification was performed in

a total volume of 25 μL in 0.2 mL reaction tubes, using a DNA thermal cycler (MyCycler, Biorad Lab, Australia). The thermal PCR profile was initial denaturation at 94 °C for 3 min, followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 53 °C for 30 s, and extension at 72 °C for 90 s and followed by final extension at 72 °C for 10 min. PCR-generated products were cloned using pGEM®-T Easy Cloning Kit (Promega, Madison, WI, USA) and the *Escherichia coli* DH5α vector as described by Vishwakarma et al. (2010). Positive insert containing amplified PCR product of each single clone (10 μL) was digested in the PCR reaction buffer with 1 U of *Alu* I and *Hae* III (New England Biolabs, Beverly, MA) (Wright and Pimm 2003) each and incubated overnight at 37 °C. The clones thus, obtained were subjected to RFLP (restriction fragment length polymorphism) analyses to generate restriction patterns. On the basis of banding pattern, operational taxonomic units (OTUs) were prepared by the presence (1)/ absence (0) of bands on the gel, and dendrogram was constructed using NTSYS software (version 2.2Windows© 2002, Applied Biostatistic Inc., Australia). At least one representative clone or two clones from each OTU were sequenced. Close relatives and phylogenetic affiliations of the sequences were determined using the BLAST search program available at the National Center for Biotechnology Information (NCBI) web site. A phylogenetic tree was constructed by 1000-fold bootstrap analysis using the neighbor-joining method, ClustalW program and NJ plot software (Higgins et al. 1994).

3.2.6 Nucleotide Sequence Accession Numbers

The clone sequences obtained in the present study were deposited with the NCBI GenBank database under the accession numbers JQ307365 to JQ307384 and JX014526 to JX014537.

3.2.7 Diversity Indices

The diversity and evenness of methanogenic community are based on the proportion of OTUs in each clone library and are calculated as Shannon-Wiener's index for diversity (Odum 1971):

$$H' = -\sum_{i=1}^{s} P_i \ln P_i \tag{3.5}$$

where s is the number of OTUs and Pi is the relative abundance of clones assigned to the i^{th} OTU to the total clones analyzed.

3.3 Results and Discussion

3.3.1 Methane Production Potential of Soil

Present study revealed variations in methane production potential across different soils. The extent of methane production ranged from 154.71 to 410.47 μg CH_4 g dws $^{-1}$ in the three soil types (Fig. 3.1). Methane production potential of Alfisol, Inceptisol, and Vertisol soils following 480 h of incubation was 154.71, 202.12, and 410.47 μg CH_4 g dws^{-1}, respectively. Methane production was the least for Alfisol and highest for the Vertisol soil. No significant difference in CH_4 production was observed between Alfisol and Inceptisol. Wang et al. (1993) reported 518 and 528 μg g^{-1} of methane from Indian paddy fields of Ludhiana and Cuttack, respectively, following 39 days of incubation. Ramakrishnan et al. (1998) reported the magnitude of methane production in alluvial and acid sulfate saline soils from 0.1 to 437.05 μg g^{-1} following 30 days of incubation of soil samples from Indian paddy fields. The study in the Philippines reported up to 40 μg CH_4 g dws^{-1} d^{-1} after laboratory incubation of 56 days (Wassmann et al. 1998). Jia et al. (2006) reported up to 140 μg CH_4 kg dws^{-1} h^{-1} for Chinese soils. The variations among such values with those of the present study may be attributed to various factors including climatic conditions, soil properties, methanogenic community, etc.

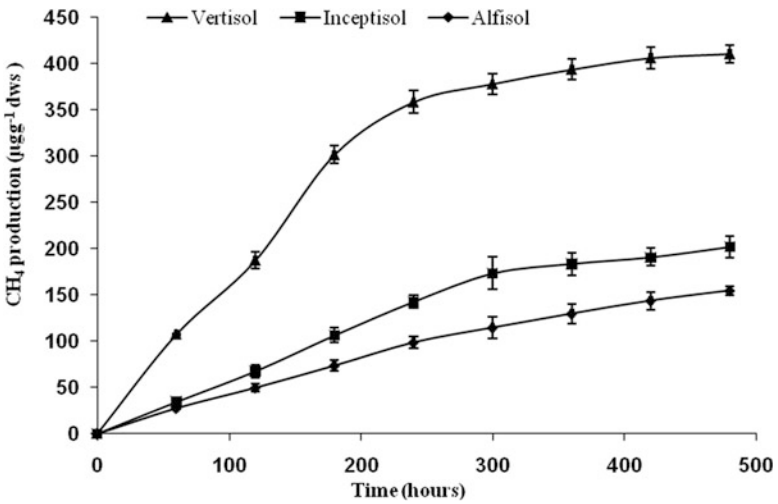


Fig. 3.1 Variation in CH_4 production potential of different soils

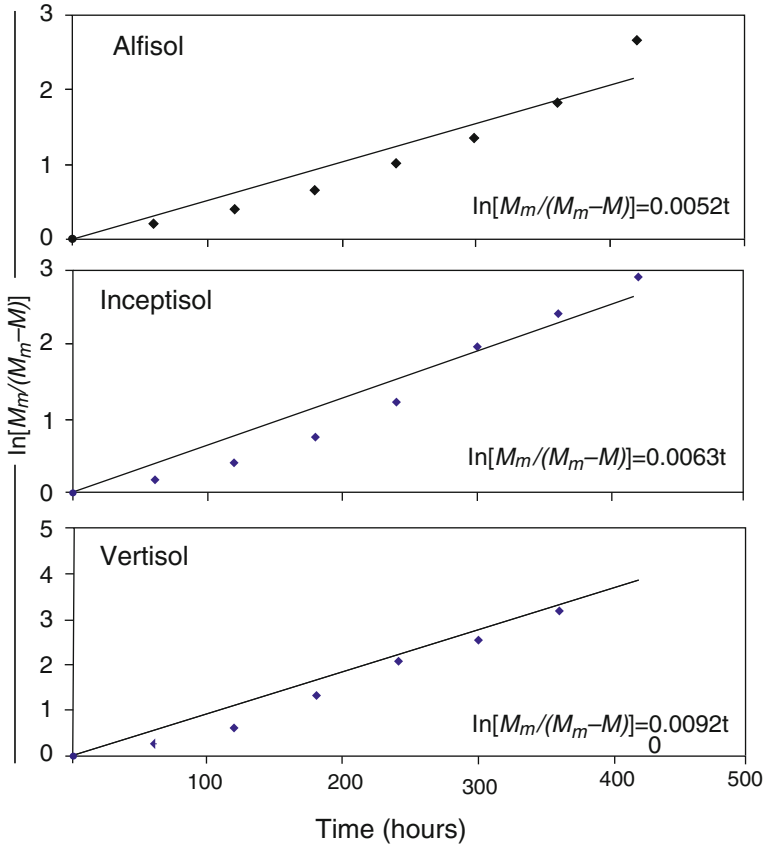


Fig. 3.2 First-order kinetic model plot for cumulative methane production in three soil types

3.3.2 Kinetics of Methane Production

The rate constant (K) values obtained from the linear plot (Fig. 3.2) for the soils studied are Alfisol, 0.0052 h^{-1}, $R^2 = 0.91$; Inceptisol, 0.0063 h^{-1}, $R^2 = 0.94$; and Vertisol, 0.0092 h^{-1}, $R^2 = 0.94$. The order of increase in K values is Alfisol < Inceptisol < Vertisol. It has been shown earlier that methane production in paddy field soil follows a first-order kinetics (Singh et al. 2012) expression given by:

$$\ln\left(\frac{M_m}{M_m - M}\right) = Kt \tag{3.6}$$

where M is methane produced at time t and M_m is the maximum amount of methane produced.

A plot of $\ln\left(\frac{M_m}{M_m - M}\right)$ versus t gives a straight line. The data on methane production plotted in Fig. 3.2 revealed that all the three soils follow first-order kinetics. It is also evident that K value for Vertisol was the highest followed by Inceptisol and Alfisol, respectively.

3.3.3 Methanogenic Archaea Community Structure

Clone libraries of methanogenic-specific 16S rRNA genes were constructed for three distinct soils, *viz.*, Alfisol, Inceptisol, and Vertisol. The RFLP pattern of 48 positive clones (screened through M13 PCR) from each soil type showed differences in the composition of methanogenic community among the three soils (Fig. 3.3). Each OTU is represented by distinct number of clone in each soil thus, contributing to the relative abundance of methanogens. The sequence analysis of distinct OTUs from all the soil samples revealed five groups of methanogens belonging to *Methanomicrobiaceae*, *Methanosarcinaceae*, *Methanosaetaceae*, *Methanobacteriaceae*, and *Methanocellales* (RC I) with their varied relative abundance in different soils (Fig. 3.3). In addition, two other non-methanogenic archaeal groups, RC II and RC IV, were also detected.

Maximum number of OTUs (S1, S2, S3, S4, S12, S13, AR1, AR2, AR3, AR5, AR6, and AR9) in Vertisol soil had resemblance with the representatives of *Methanomicrobiaceae*, *Methanosarcinaceae*, *Methanosaetaceae*, and RC I (Fig. 3.3). The clones from Inceptisol (S5, S6, S9, S10, S11, S14, AS7, AS8, AR7, AR8, and AR10) and Alfisol (S15, S16, S17, S18, S19, S20, S21, S22, and AR11) soils were closest to RC I, *Methanomicrobiaceae*, *Methanobacteriaceae*, *Methanosarcinaceae* and RC I, *Methanomicrobiaceae*, and *Methanosaetaceae*, respectively (Fig. 3.3). Various cultured methanogenic groups also exhibited close similarity with the clones isolated presently. The clones S16 and S19 from Alfisol soil showed close similarity (96%) with *Methanogenium thermophilum*. The clone AR1 of Vertisol soil was 98% closely related to *Methanospirillum hungatei*. AR 10 of Inceptisol is 98% closely related to *Methanosarcina barkeri*.

Figure 3.4 shows the relative proportion of different methanogenic groups in the respective soil types with highest percentage of RC I in Alfisol (55.75%) followed by Vertisol (52.97%) and Inceptisol (34.58%). *Methanomicrobiaceae* was also common among all the groups with the abundance of 43.75% in Inceptisol, 30.04% in Alfisol, and 27.03% in Vertisol. *Methanosaetaceae* was found only in Vertisol and Alfisol with 9.28% and 14.21%, respectively. *Methanosarcinaceae* was detected in Vertisol (10.72%) and Inceptisol (15.67%) but not in Alfisol. *Methanobacteriaceae* is the group which was only detected in the Inceptisol with 6% abundance. Thus, all the soils had their distinct population group of methanogenic community.

The above results indicate that the diversity of methanogens and relative abundance varied with the soil type. These methanogenic groups belong to two main categories based on the availability of substrates, i.e., hydrogenotrophic (RC I,

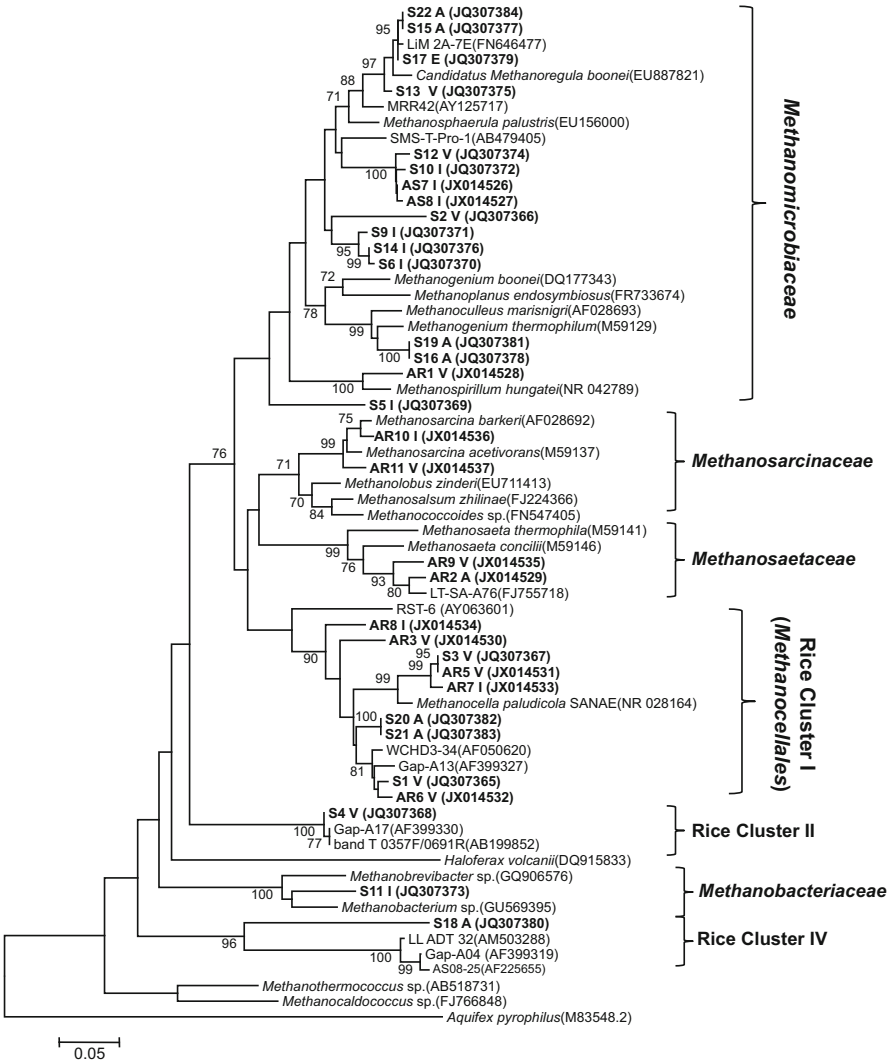


Fig. 3.3 Phylogenetic relationship of methanogenic archaeal 16S rRNA genes retrieved from cloning-RFLP-sequencing analysis in three types of soil, i.e., Alfisol (A), Inceptisol (I), and Vertisol (V). Nucleotide sequences obtained from this study are presented in bold letters. Scale bars represent an estimated 5% sequence divergence. Above 70% bootstrap values are shown at each branch point. *Aquifex pyrophilus* was used as an out-group. GenBank accession numbers are indicated for each sequenced clone in parenthesis

Methanomicrobiaceae, *Methanobacteriaceae*, and *Methanosarcinaceae*) and acetotrophic (*Methanosarcinaceae* and *Methanosaetaceae*). The RC I group of methanogens was dominant in all soil types (Fig. 3.4). This is in accordance of our earlier studies (Dubey et al. 2013) where diversity analysis by DGGE reveals the

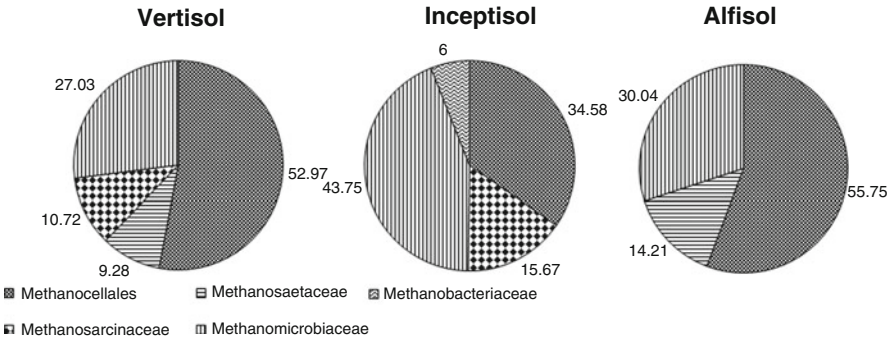


Fig. 3.4 Relative proportion of methanogenic archaeal community across the different soil types

same dominant group. This is also in line with other reports (Ramakrishnan et al. 2001; Hashimoto et al. 2005) on the rhizospheric methanogens common in different soils. According to Conrad et al. (2006), RC I prefers extreme conditions (acidic pH or higher temperature) or very low H_2 concentration. Thus, compared to other soils under the prevailing acidic conditions (pH 6.2), the dominance of RC I group (with 55.8% abundance) has been observed in the methanogenic archaeal groups. *Methanomicrobiaceae* is the other hydrogenotrophic group, inhabiting rice roots, that utilizes H_2-CO_2, and formate as the energy source was found in all the soil types studied. This group was also detected in soils from various geographic locations (Ramakrishnan et al. 2001; Watanabe et al. 2006; Conrad et al. 2008). This shows the ubiquity of *Methanomicrobiaceae* in every climate regime because of their ability to grow even at low H_2 concentrations up to a certain limit. However, the relative abundance indicates that these were less dominant in rice root niche compared to RC I in all soils (Fig. 3.4). This study revealed lesser number of *Methanobacteriales* compared to other root-inhabiting groups, i.e., only 6% in Inceptisol soil. According to Lu and Conrad (2005), *Methanobacteriales* are found in the rice roots, and their growth is favored by high H_2 concentration. *Methanobacteriales* showed low intensity compared to other methanogenic groups, RC I, and *Methanomicrobiales* in rice roots (Scheid et al. 2003). Similar results were found in our previous study (Dubey et al. 2014). However, *Methanomicrobiaceae* was not detected in Vertisol and Alfisol soil. This could be due to the use of the different technique, DGGE, for diversity analysis in the previous study which is unable to detect the community representing <1% of the population (Murray et al. 1998).

The present study reveals that Vertisol soil inhabited more diverse group of methanogenic archaea and Alfisol was found to harbor least diverse group. The Shannon-Wiener diversity index (H') values of 1.17, 1.01, and 0.97 for Vertisol, Inceptisol, and Alfisol soils, respectively, for the methanogen community also support this finding.

The members of *Methanosarcinaceae* and *Methanosaetaceae* were found together in Vertisol soil, while Inceptisol soil contained *Methanosarcinaceae* and

the Alfisol only *Methanosaetaceae*. Thus, one can infer that Alfisol soil has acetate concentration relatively lower for methanogenesis and permits only the *Methanosaeta* population to proliferate compared to *Methanosarcina* that grow only at higher acetate concentrations (Kruger et al. 2005). *Methanosarcinaceae* is relatively versatile in substrate utilization and can use acetate, methanol, and H_2-CO_2 unlike the strict dependence of *Methanosaetaceae* only on acetate (Conrad 2007). Thus availability of acetate might have differentiated the acetotrophic community structure in the soil from the three studied fields.

All the soil types harbor both acetotrotrophic and hydrogenotrophic methanogens, but the community structure variation indicates their differential participation in energy and C-flow in the immediate soil environment that ends up with the pronounced changes in methane production. In addition, the high spatial variability of organic carbon in different soils may also be responsible for variations in microbiological processes and the distribution of methanogenic community (Wachinger et al. 2000).

3.3.4 Soil Physicochemical Properties

Table 3.1 shows the variation in properties of soils measured in this study. The range of pH varied among three types of soils and found lowest in Alfisol which is acidic. Highest is found in Vertisol soil. Redox potential (E_h) values are found with little variation and with negative E_h. Higher organic C content is found in Vertisol followed by Inceptisol and Alfisol. NH_4^+-N content varied among all soil types.

Soil physicochemical properties can affect microbial community either directly by providing specific habitat for selecting microbial community or indirectly effecting root exudation that influence the CH_4 production potential of soils (Yao et al. 1999; Mitra et al. 2002; Teh and Silver 2006). The methanogens are pH-sensitive and grow within the narrow pH range of 6–8 (Chang and Yang 2003). Neutral pH favors methanogenesis (Wang et al. 1993). Comparatively, pH of Alfisol soil was acidic and likely to reduce the activity of methanogens with the ultimate low methane output. It has been observed that hydrogenotrophs outcompete the acetotrophs in the Alfisol soils, and only about 14% abundance of acetotrophic community detected presently was the lowest among all the soil types. In acidic pH, the unionized fatty acids accumulate to lower the pH and also inhibit methanogenesis from acetate (Beccari et al. 1996). The highest pH value of Vertisol accompanied by the highest methane production could be attributed to the rise in the population of the potent methane-generating archaea, i.e., acetotrophic methanogens, that grow almost exponentially with the rise in pH (Kravchenko and Yu 2006). The E_h (redox potential) values ranged from −92.2 to −60.5 mV (Table 3.1) which indicate the reduced state of soil, a prerequisite for methanogenesis. Almost similar results from laboratory incubation study were obtained by Yao et al. (1999) as the E_h for methane production in different paddy soils ranged from −135 to 30 mV. The E_h values of the present study are quite high

compared to the range reported earlier, i.e., −150 mV (Wang et al. 1993) and < −200 mV (Mitra et al. 2002). Gaunt et al. (1997) also reported methane production at higher E_h. Hence, the present study infers that there exists an anaerobic environment that promotes methanogenesis even at higher E_h.

It is well established that soil organic carbon and total nitrogen have significant effect methanogenesis (Yao et al. 1999; Liu et al. 2011). High organic matter content favors the intensity of reductive soil process (Neue et al. 1996). Therefore, relatively higher organic C in the Vertisol soil (Table 3.1) could have provided the favorable environment for methanogens, while Alfisol soil with comparatively low organic C could limit their existence. Higher diversity in Vertisol could also be due to high NH_4^+-N content. A nitrogen fertilizer like urea in the soil yields NH_4^+ ions on hydrolysis and favors methanogen existence by maintaining optimal pH or decreasing the redox potential (Wang et al. 1992). Our earlier studies also showed a positive correlation between NH_4^+-N and methanogenesis (Dubey et al. 2013, 2014).

3.4 Conclusions

Thus, it is inferred that variations in soil physical and chemical characteristics affected the methane production potential of the soils studied with minor alteration in methanogenic archaeal community composition.

3.5 Future Perspectives

The present study establishes that the soil characteristics control the methane emission and methanogen community structure. It would be further interesting and useful to prepare a database for methanogens and methane production in different soil types to come out with a generalized model for methane production from paddy fields across the country. The analysis of methanogenic community structure and function in relation to soil type will be helpful in enhancing the understanding of microbial ecology of methanogens and methane production potential of soils that is essential for formulation of strategies to mitigate methane emission from rice fields.

Acknowledgments This study was financially supported by Department of Science and Technology, Government of India, New Delhi.

References

Adhya TK, Bharati K, Mohanty SR et al (2000) Methane emission from rice fields at Cuttack, India. Nutr Cycl Agroecosyst 58:95–105

Anand S, Dahiya RP, Talyan V et al (2005) Investigations of methane emissions from rice cultivation in Indian context. Environ Int 31:469–482

Anastasi C, Dowding M, Simpson VJ (1992) Future CH_4 emissions from rice production. J Geophys Res 97:7521–7752

APHA American Public Health Association (1985) Standard methods for the examination of water and wastewater. American Public Health association, Washington, DC

Aulakh MS, Wassmann R, Bueno C et al (2001) Impact of root exudates of different cultivars and plant development stages of rice (*Oryza sativa* L.) on methane production in paddy soil. Plant Soil 239:77–85

Beccari M, Bonemazzi F, Majone M et al (1996) Interaction between acidogenesis and methanogenesis in the anaerobic treatment of olive oil mill effluents. Water Res 30:183–189

Blake DR, Rowland FS (1988) Continuing worldwide increase in tropospheric methane, 1978 to 1987. Science 239:1129–1131

Borrel G, Parisot N, Harris HMB et al (2014) Comparative genomics highlights the unique biology of *Methanomassiliicoccales*, a Thermoplasmatales-related seventh order of methanogenic archaea that encodes pyrrolysine. BMC Genomics 15:679

Bouwman AF (1990) Soils and the greenhouse effect. Wiley, Chichester

Cao P, Zhang LM, Shen JP et al (2012) Distribution and diversity of archaeal communities in selected Chinese soils. FEMS Microbiol Ecol 80:146–158

Chang TC, Yang SS (2003) Methane emission from wetlands in Taiwan. Atmos Environ 37:4551–4558

Chen YH, Prinn RG (2006) Estimation of atmospheric methane emissions between 1996 and 2001 using three dimensional chemical transport model. J Geophys Res 111:1–25

Chin KJ, Lukow T, Conrad R (1999) Effect of temperature on structure and function of the methanogenic archaeal community in an anoxic rice field soil. Appl Environ Microbiol 65:2341–2349

Conrad R (1996) Soil microorganisms as controllers of atmospheric trace gases (H_2, CO, CH_4, OCS, N_2O and NO). Microbiol Rev 60:609–640

Conrad R (2007) Microbial ecology of methanogens and methanotrophs. Adv Agron 96:1–63

Conrad R, Erkel C, Liesack W (2006) Rice cluster I methanogens, an important group of archaea producing greenhouse gas in soil. Curr Opin Biotechnol 17:262–267

Conrad R, Klose M, Noll M et al (2008) Soil types links microbial colonization of rice roots to methane emission. Glob Chang Biol 14:657–669

Conrad R, Klose M, Lu Y et al (2012) Methanogenic pathways and archaeal communities in three different anoxic soils amended with rice straw and maize straw. Front Microbiol 3:1–12

Drake HL (1994) Acetogenesis. Chapman and Hall, New York

Dubey SK, Singh A, Singh RS et al (2013) Changes in methanogenic population size and CH_4 production potential in response to crop phenology in tropical rice field. Soil Biol Biochem 57:972–978

Dubey SK, Singh A, Watanabe T et al (2014) Methane production potential and methanogenic archaeal community structure in tropical irrigated Indian paddy soils. Biol Fertil Soils 50:369–379

Garcia JL (1990) Taxonomy and ecology of methanogens. FEMS Microbiol Rev 87:297–308

Gaunt JL, Grant IF, Nee HU et al (1997) Soil characteristics that regulate soil reduction and methane production in wetland rice soils. Soil Sci Soc Am J 61:1526–1531

Gupta PK, Gupta V, Sharma C et al (2009) Development of methane emission factors for Indian paddy fields and estimation of national methane budget. Chemosphere 79:590–598

Hashimoto YT, Ikenga M, Asakawa S et al (2005) Effect of free-air CO_2 enrichment (FACE) on methanogenic archaeal communities inhabiting rice roots in Japanese rice field. Soil Sci Plant Nutr 51:91–100

Higgins D, Thompson J, Gibson T et al (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Hoshino YT, Morimoto S, Hayatsu M et al (2011) Effect of soil type and fertilizer management on archaeal community in upland field soils. Microbes Environ 307:16–26

Hou AX, Wang ZP, Chen GX et al (2000) Effects of organic and N fertilizers on methane production in a Chinese rice soil and its microbiological aspect. Nutr Cycl Agroecosyst 58:333–338

IPCC Intergovernmental Panel on Climate Change (1995) Climate change 1994 (radiative forcing of climate change). Cambridge University Press, Cambridge

IPCC Intergovernmental Panel on Climate Change (1996) XII summary for policy makers. In: Houghton IT, Meira F, Callander LG, Harris BA, Kattenberg A, Maskell K (eds) Climate change 1995: the scientific basis of climate change. Cambridge University Press, Cambridge, p 572

IPCC International Panel on Climate Change (2013) Climate change. The physical science basis. Oxford Press, Cambridge, p 09

Jackson ML (1958) Soil chemical analysis. Prentice-Hall, Englewood Cliffs

Jia Z, Cai Z, Tsuruta H (2006) Effect of rice cultivar on CH_4 production potential of rice soil and CH_4 emission in a pot experiment. Soil Sci Plant Nutr 52:341–348

Kniffin JK, Freyre D, Balser TC (2010) Methane dynamics across wetlands plant species. Aquat Bot 93:107–113

Kravchenko IK, Yu K (2006) Relationship between major soil properties and culturable microorganisms affecting CH_4 and N_2O dynamics in rice soils. Arch Agron Soil Sci 52:607–615

Kruger M, Frenzel P, Kemnitz D et al (2005) Activity, structure and dynamics of the methanogenic archaeal community in a flooded Italian rice field. FEMS Microbiol Ecol 51:323–331

Kumarswamy S, Ramakrishnan B, Setunathan N (2001) Methane production and oxidation in an anoxic rice soil as influenced by inorganic redox species. J Environ Qual 30:2195–2201

le Mer J, Roger P (2001) Production, oxidation, emission and consumption of methane by soils: a review. Eur J Soil Biol 37:25–50

Liu DY, Ding WX, Jia ZJ et al (2011) Relation between methanogenic archaea and methane production potential in selected natural wetland ecosystems across China. Biogeosciences 8:329–338

Lu YH, Conrad R (2005) In situ stable isotope probing of methanogenic archaea in rice rhizosphere. Science 309:1088–1090

Lu Y, Murase J, Watanabe A et al (2004) Linking microbial community dynamics to rhizosphere carbon flow in a wetland rice soil. FEMS Microbiol Ecol 48:179–186

Ludmila C, Julia A, Rudoyk R et al (1998) C1 transfer enzymes and coenzymes linking methylotrophic bacteria and methanogenic archaea. Science 281:99–101

Mitra S, Wassmann R, Jain MC et al (2002) Properties of rice soils affecting methane production potentials: 1. Temporal patterns and diagnostic procedures. Nutr Cycl Agroecosyst 64:169–182

Murray AE, Preston CM, Massana R et al (1998) Seasonal and spatial variability of bacterial and archaeal assemblages in the coastal waters near Anvers Island, Antarctica. Appl Environ Microbiol 64:2585–2595

Neue HU, Sass RL (1994) Trace gas emissions from rice soils. In: Prinn RG (ed) Global atmospheric-biospheric chemistry. Plenum Press, New York, pp 119–148

Neue HU, Wassmann R, Lantin RS et al (1996) Factors affecting methane emission from rice fields. Atmos Environ 30:1751–1754

Nguyen C (2003) Rhizo-deposition of organic C by plants: mechanism and controls. Agronomie 23:375–396

Odum EP (1971) Fundamentals of ecology. Saunders, Philadelphia

Pathak H, Prasad S, Bhatia A et al (2003) Methane emission from rice-wheat cropping system in the Indo-Gangetic plain in relation to irrigation, farmyard manure and dicyandiamide application. Agric Ecosyst Environ 97:309–316

Prather M, Ehhalt D (2001) Atmospheric chemistry and greenhouse gases. In: Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell KJ (eds) Climate change 2001: the scientific basis, contribution of working group I to the third assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, pp 239–287

Ramakrishnan B, Kumaraswamy S, Mallick K et al (1998) Effect of various anionic species on net methane production in flooded rice soils. World J Microb Biotechnol 14:743–749

Ramakrishnan B, Lueders T, Dunfield PF et al (2001) Archaeal community structures in rice soils from different geographical regions before and after initiation of methane production. FEMS Microbiol Ecol 37:175–186

Rath AK, Ramakrishnan B, Rao VR et al (2005) Effects of rice-straw and phosphorus application on production and emission of methane from tropical rice soil. J Plant Nutr Soil Sci 168:248–254

Reeburgh WS, Whjalen SC, Alperin MJ (1993) The role of methylotrophy in the global methane budget. In: Murrell JC, Kelly DP (eds) Microbial growth on CI compound. Intercept Ltd, Andover, pp 1–14

Rodhe H (1990) A comparison of the contribution of various gases to the greenhouse effect. Science 248:1217–1219

Sass RL, Fisher FM, Lewis ST et al (1994) Methane emissions from rice fields: effect of soil properties. Global Biogeochem Cycles 8:135–140

Scheid D, Stubner S, Conrad R (2003) Effects of nitrate- and sulfate-amendments on methanogenic populations in rice root incubations. FEMS Microbiol Ecol 43:309–315

Singh S, Singh JS, Kashyap AK (1999) Methane flux from irrigated rice fields in relation to crop growth and N-fertilization. Soil Biol Biochem 31:1219–1228

Singh A, Singh RS, Upadhyay SN et al (2012) Community structure of methanogenic archaea and methane production associated with compost-treated tropical rice-field soil. FEMS Microbiol Ecol 82:118–134

Teh YA, Silver WL (2006) Effects of soil structure destruction on methane production and carbon partitioning between methanogenic pathways in tropical rain forest soils. J Geophys Res 111:1–8

Uprety DC, Baruah KK, Borah L (2011) Methane in rice agriculture: a review. J Sci Ind Res 70:401–411

USDA FAS United States Department of Agriculture Foreign Agricultural Service (2009) World agricultural production. Office of Global Analysis, Circular Series, WAP, pp 07–09

Vishwakarma P, Singh M, Dubey SK (2010) Changes in methanotrophic community composition after rice crop harvest in tropical soils. Biol Fertil Soils 46:471–479

Wachinger G, Fiedler S, Zepp K et al (2000) Variability of soil methane production on the micro-scale: spatial association with hot spots of organic material and archaeal populations. Soil Biol Biochem 32:1121–1130

Wagner D, Pfeiffer EM, Bock E (1999) Methane production in aerated marshland and model soils: effects of microflora and soil texture. Soil Biol Biochem 31:999–1006

Walkley A (1947) A critical examination of a rapid method for determining organic carbon in soils-effects of variations in digestion conditions and of inorganic soil constituents. Soil Sci 63:251–264

Wang ZP, Lindau CW, de Laune RD et al (1992) Methane production from anaerobic soil amended with rice straw and nitrogen fertilizers. Fertil Res 33:115–121

Wang ZP, Lindau CW, de Laune RD et al (1993) Methane emission and entrapment in flooded rice soils as affected by soil properties. Biol Fertil Soils 16:163–168

Wang G, Watanabe T, Jin J et al (2010) Methanogenic archaeal communities in paddy field soils in north-east China as evaluated by PCR-DGGE, sequencing and real-time PCR analyses. Soil Sci Plant Nutr 56:831–838

Wassmann R, Neue HU, Bueno C et al (1998) Methane production capacities of different rice soils derived from inherent and exogenous substrates. Plant Soil 203:227–237

Watanabe T, Asakawa S, Nakamura A et al (2004) DGGE method for analyzing 16S rDNA of methanogenic archaeal community in paddy field soil. FEMS Microbiol Lett 232:153–163

Watanabe T, Kimura M, Asakawa S (2006) Community structure of methanogenic archaea in paddy field soil under double cropping (rice-wheat). Soil Biol Biochem 38:1264–1274

Watanabe T, Wang G, Lee CG et al (2011) Assimilation of glucose-derived carbon into methanogenic archaea in soil under unflooded condition. Appl Soil Ecol 48:201–209

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci 87:4576–4579

Wright AGD, Pimm C (2003) Improved strategy for presumptive identification of methanogens using 16S ribotyping. J Microbiol Methods 55:337–349

Yan X, Ohara T, Akimoto H (2003) Development of region-specific emission factors and estimation of methane emission from rice fields in east, southeast and South Asian countries. Glob Chang Biol 9:237–254

Yao H, Conrad R, Wassmann R et al (1999) Effect of soil characteristics of sequential reduction and methane production in sixteen rice paddy soils from China, Philippines and Italy. Biogeochemistry 4:269–295

Zinder SH (1993) Physiological ecology of methanogens. In: Ferry JG (ed) Methanogenesis. Chapman Hall, New York, pp 128–206

Chapter 4
In Silico Study of the Geminiviruses Infecting Ornamental Plants

Avinash Marwal, Megha Mishra, Rakesh Verma, Rajneesh Prajapat, and R. K. Gaur

Abstract Over the past few decades, there has been more interest in Geminiviruses, especially Mastrevirus and Begomovirus, as many of the diseases they cause have now reached epidemic magnitude. Ornamental plants are widely distributed in India and across the globe having high environmental adaptability. Their farming forms a major branch of horticulture. At most of the places, crops stay in the field for a particular season, while different ornamental plants grow in or nearby these agricultural fields throughout the year. Ornamental plants serve as an alternative host for Geminiviruses in the absence of the main crops and considered as a source of new viruses or reservoirs of unidentified viruses which are often neglected during diversity studies. Ornamental plants may allow the spread and transmission of Geminiviruses back to crop plants when the cropping season returns, which enhances the host range of these viruses. Thus, there is a pressing need for additional information on the diversity and distribution of Geminiviruses in ornamental plants.

Keywords Geminiviruses · Mastrevirus · Begomovirus · Ornamental plants · Alternative host

4.1 Introduction

Ornamental plants forms one of the major branches of horticulture and are grown for decoration of gardens and designing landscape for the purpose of enjoyment of gardeners, visitors, and the public (Mahesh et al. 2012). They are also used for beautification of homes (in house plants), as cut flowers and specimen display (Prajapat et al. 2012a; Marwal et al. 2013a). Majorly ornamental garden plants are grown for its aesthetic appearance such as flowers, scent, leaves, fruit, stem, bark, overall foliage texture, and aesthetic form and characteristics (Mahesh et al. 2012; Marwal et al. 2013b). Likewise certain trees may be called ornamental trees when

A. Marwal · M. Mishra · R. Verma · R. Prajapat · R. K. Gaur (✉)
Department of Biosciences, College of Arts, Science and Humanities, Mody University, Sikar, Rajasthan, India

D. K. Choudhary et al. (eds.), *In Silico Approach for Sustainable Agriculture*,
https://doi.org/10.1007/978-981-13-0347-0_4

they are used as part of landscape setting (Khan et al. 2005). The quality and quantity of ornamental plants are ascertained by biotic agents like plant viruses which are an important bottleneck in the production (Marwal et al. 2014). Among plant viruses, *Geminiviridae* is the largest family infecting a number of crops, weeds, and ornamental plants (Prajapat et al. 2011; Marwal et al. 2013c). They belong to class II of viruses and the largest single-stranded DNA viruses' family. Members of this family infect both monocotyledonous and dicotyledonous plants. There are at present more than 300 species in this family which are divided among seven well-characterized genera (Mastrevirus, Begomovirus, Curtovirus, Topocuvirus, Eragrovirus, Becurtovirus, and Turncurtovirus) (Fauquet et al. 2011; Hernandez-Zepeda et al. 2013).

Disease symptoms caused by Geminiviruses are yellowing of leaf veins, bright yellow mosaic patterns, yellow mottle leaves, curling of leaves, stunted growth, leaf distortion, and leaf streaks (Rybicki and Hughes 1990; Konate et al. 1995; Prajapat et al. 2012b; Gilbertson et al. 1993; Thomas et al. 1986; Navot et al. 1991). These diseases are responsible for severe threat to the food production in the developing countries residing in the tropical and subtropical regions of the world (Moffat 1999) due to a number of reasons, such as recombination between Geminiviruses, transfer of diseased plant material to a new place, agriculture practices into new cultivated region, and migration of vectors that can spread the virus from one plant to another (Gilbertson et al. 1991; Marwal et al. 2012). Mastreviruses transmission is caused by the vector leafhoppers (*Cicadulina mbila*) (Muhire et al. 2013), Begomoviruses are transmitted by the vector whiteflies (*Bemisia tabaci*), Topocuviruses and Curtoviruses are transmitted by the vector treehoppers (*Micrutalis malleifera*), and Eragroviruses, Becurtoviruses, and Turncurtoviruses are transmitted by the vector leafhoppers (*Circulifer haematoceps*) (Varsani et al. 2014; Soto et al. 2005).

Mastreviruses comprises of a single genome component (ssDNA) which has the capability of infecting both monocotyledonous and dicotyledonous plants and prevails in the Old World. Curtoviruses and Topocuviruses also have one genomic DNA and infect only dicots in the New World (Hernandez-Zepeda et al. 2013; Briddon et al. 1996). Begomoviruses comprises the largest genera and are found in the Old and New World. Begomoviruses have either monopartite genomes or bipartite genomes known as DNA-A and DNA-B. Sometimes it has been found that a number of monopartite Begomoviruses are associated with satellite molecules (alphasatellites and betasatellites) (Hanley-Bowdoin et al. 2013). Eragroviruses and Becurtoviruses comprise of a single genome infecting only dicots in the Old World (Heydarnejad et al. 2007; Varsani et al. 2009a, b; Baliji et al. 2004). Ornamentals play an important role in persistence and spread of Geminiviruses which they serve as a potential alternative host of their primary inoculums. A scan through literature revealed that there is growing information on Geminiviruses infecting ornamental plants and investigations are therefore taken up with the aim of assessing the status of disease in India and at the global level for understanding the causal nature and severity of the vector pathogens. This chapter thus presents a tip of the iceberg of the diversity of Geminiviruses infecting ornamentals in India and across the globe.

4.1.1 Old World Echelon of Geminiviruses Infecting Ornamental Plants

Virus diseases of Geminiviruses are characterized with a number of visible symptoms. These diseases may lead to noteworthy yield losses to ornamental plants in field circumstances render them unfit for market use (Garcia-Arenal et al. 2001; Herrera-Vásquez et al. 2013). Detail on the distribution, incidence, and occurrence and of the Geminiviruses is necessitated in order to build up strategies for disease management. Similarly, one of the significant features in management of viral diseases is to investigate and comprehend of its survival in alternative hosts (Lapido 1988; Pruss et al. 1997; Weinberger and Msuya 2004). Geminiviruses in alternative hosts may serve as the source inoculum for fresh crops even when recognized by using virus-free seeds. Given below are the details of the diversity of various ornamental plants in the Old World serving as an alternate host of Geminiviruses.

4.1.2 Asia

In recent years the *Geminiviridae* family has received a great deal of attention, and it is one of the most important studied plant virus families. Several reviews on Geminiviruses covered different aspects of their serological properties, epidemiology, biology, and molecular biology (Polston and Anderson 1997; Hanley-Bowdoin et al. 1999; Morales and Anderson 2001). A new disease of Geminivirus *Tobacco leaf curl Japan virus* isolated from an ornamental plant *Lonicera japonica* (Honeysuckle) (Osaki and Inouye 1983) was reported in Japan exhibiting veinal chlorosis symptoms. The *Tobacco leaf curl Japan virus* genome showed a highest identity of 92.94% with previously reported *Tobacco leaf curl Japan virus* (Ali et al. 2014). Further in an another report from India, ornamental species of Jatropha, viz., *J. integerrima*, *J. multifida*, and *J. podagrica*, which are also used as biofuel (Speight and Singh 2014; Openshaw 2000) were found with a Geminivirus infection (mosaic symptoms) in CSIR-NBRI garden of Lucknow, India, characterized with molecular techniques. Based on the highest sequence identities of partial DNA-A genome (~1.2 kb) and close phylogenetic relationships, *Tomato leaf curl Patna virus* in *J. multifida* (HQ848381), *Jatropha mosaic India virus* in *J. podagrica* (HQ848382), and *Papaya leaf curl virus* in *J. integerrima* (JQ043440) were identified (Snehi et al. 2016).

Similarly Marwal et al. (2013a) identified a Begomovirus, which was found associated with its satellites molecules infecting a new host *Tagetes patula* in India. The identified virus reveals closeness with other same isolates of its family (Fig. 4.1). Further complete nucleotide was isolated from 15 samples having disease symptoms revealed positive hybridization results to a radioactively labeled probe of DNA-A of *Tomato leaf curl New Delhi virus*, confirming the prevalence of a Begomovirus in *Clerodendron inerme* (family Verbenaceae) (Chethana et al.

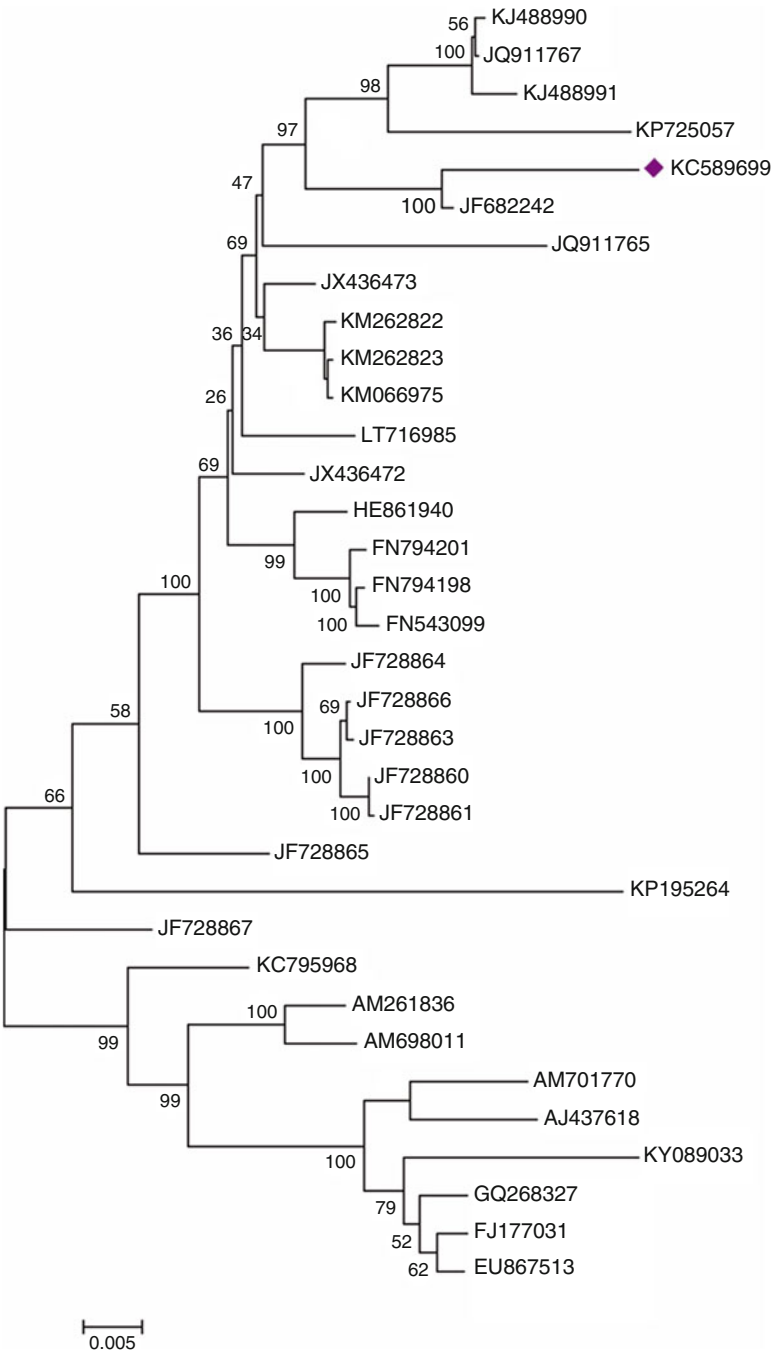


Fig. 4.1 The neighbor joining (NJ) phylogenetic tree based upon the alignments of the sequences of *Ageratum enation virus* with other geographical isolates of *Ageratum enation virus*

2013), a common ornamental plant, grown as a hedge in the gardens of India. Typical symptoms observed were bright yellow mosaic spots, small leaves, and stunted growth. Sequence analysis with sequences of other Begomoviruses showed the Geminivirus most similar to *Cotton leaf curl Kokhran virus* and *Tomato leaf curl New Delhi virus* (John et al. 2006). In an another incidence, Croton, also known as *Codiaeum variegatum* L., is cultivated as a famous ornamental plant (Mahesh et al. 2008) in the home gardens, as well garden of botanical importance and office gardens in and around Bengaluru, South India, which was observed with a leaf curl disease, severe inward curling, vein thickening, stunting, and reduction in leaf size. The phylogenetic analysis suggested that the identified Geminivirus was similar to many other Begomoviruses like *Croton leaf curl virus* and *Ageratum yellow vein virus* from the Indian subcontinent that infects tobacco, tomato, papaya, and cotton (Mahesh et al. 2010). Moreover in Lucknow, India, *Solanum capsicastrum* from the family Solanaceae also known as winter cherry grown in pots and beds in gardens for its ornamental value (Deepa and Jofeena 2015) was observed with a leaf curl disease accompanied by green blisters and distortions. Whiteflies transmission studies were positive on healthy plants producing symptoms similar to Geminivirus infection. Sequence analysis of the Begomovirus revealed 93% identity with *Pepper leaf curl Lahore virus* of *Capsicum annuum* from Pakistan (Srivastava et al. 2013). In an another report from Main Research Station (MRS), Raichur, Karnataka, Southern India sunflower plants were observed with a leaf curl disease and successfully transmitted by whiteflies. Symptoms like curling, malformation of leaves, leaf thickening, leaf enations, and severe stunting were observed on sunflower (Baldotto and Baldotto 2015) infected plants. Sequence analysis revealed of this Geminivirus with Begomoviruses clustered with *Tomato leaf curl Karnataka virus* (EU604297) (Govindappa et al. 2011).

In a remarkable report, first time a Begomovirus and associated betasatellite was found infecting *Rosa indica* plants in India. Its genome showed a greater diversity with other Begomoviruses reported from neighboring countries (Fig. 4.2) and poses a serious threat to other economically important crop species in India (Marwal et al. 2017; Sahu et al. 2014). *Zinnia elegans* an ornamental plant (Riaz et al. 2008) cultivated in India and during a survey were characterized with leaf curling, foliar deformation and distortion symptoms in Gorakhpur, and nearby locations of North Eastern Uttar Pradesh, India. Whiteflies were found in association with *Zinnia elegans* pointed the possible presence of Begomovirus in the field. Based on the highest nucleotide similarity of 97% and 99% at amino acid level, the closest relationship was showed with the isolates of *Zinnia leaf curl virus* (Pandey and Tiwari 2012). African daisy (*Dimorphotheca sinuate*) plant from the family Asteraceae, an ornamental plant (Uga 2005) used chiefly as a bedding plant in the garden area, was observed with yellow vein net symptoms in the year 2006 growing in the NBRI garden at Lucknow which showed nucleotide sequence of 98–97% identity with *Tomato leaf curl virus* (ToLCV) and 96% identity with *Tobacco curly shoot virus* (TobCSV) isolates (Raj et al. 2007). Similarly yellow vein net disease, shortening of leaves and stunting of plants, was observed in an ornamental from India plant belonging to family Asteraceae which is the *Calendula officinalis*

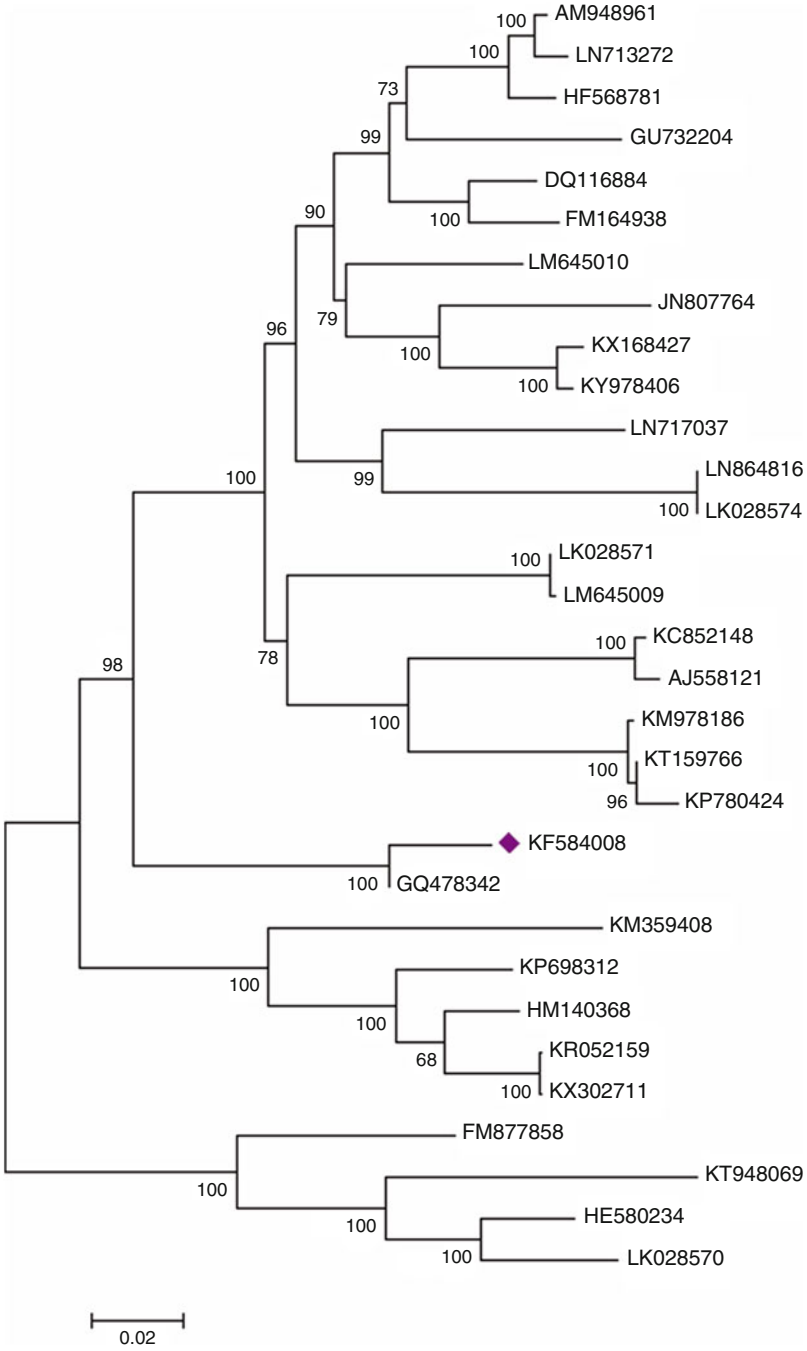


Fig. 4.2 Phylogenetic tree of *Rose leaf curl virus* with other Begomoviruses showing more than 85% sequence similarities with our isolate. Here, KF584008, *Rose leaf curl virus* isolate sikar

(Kareem et al. 2014) in the gardens of Aligarh and Lucknow. The Geminivirus was successfully transmitted from naturally infected *Calendula* to healthy seedlings by whiteflies (*Bemisia tabaci*). Sequence analysis revealed the highest nucleotide sequence identities (95%, 94%, and 93%) of the virus infecting *C. officinalis* with *Tobacco curly shoot virus* (AF240675), *Ageratum enation virus* (AJ437618), and *Tomato leaf curl Bangladesh virus* (AF188481), respectively (Khan et al. 2005).

In Guangdong, China, leaf curl disease symptoms were observed in Allamanda plants. Allamanda is an ornamental flowering plant (Chaveerach et al. 2014) also known as yellow bell, golden trumpet, or buttercup flower, and is a genus of tropical shrubs or vines belonging to the dogbane family (Apocynaceae). The Geminivirus DNA-A suggested the highest sequence identity (81.2%) to *Tomato leaf curl Guangdong virus*. The threshold value for demarcation of Begomovirus is less than 89%. Hence *Allamanda leaf curl virus* name was proposed for this new Geminivirus (He et al. 2009). Further in Jiangsu province of China, yellow mosaic symptoms were observed in *Clerodendrum cyrtophyllum* (Hamdy et al. 2007), and two Geminiviruses (YX2-I and YX2-II) were identified from the collected samples. Computational softwares suggested YX2-I to be a different species of Begomovirus named were given as *Clerodendrum golden mosaic Jiangsu virus*. Whereas YX2-II were an isolate of bipartite Begomovirus *Clerodendrum golden mosaic China virus*. In infectivity study, DNA-A of *Clerodendrum golden mosaic China virus* was capable of showing interaction with the betasatellite associated with *Tobacco curly shoot virus* producing symptoms in *Nicotiana glutinosa* and *Nicotiana benthamiana* plants. The *Clerodendrum golden mosaic Jiangsu virus* easily interacts with *Tobacco curly shoot virus* but negative with *Clerodendrum golden mosaic China virus* DNA-B in *Nicotiana benthamiana* host (Li and Zhou 2010). Similarly an

Fig. 4.2 (continued) AS22; GQ478342, *Rose leaf curl virus* clone RoLCuV-[PK,Fai,06]; HM140368, *Papaya leaf crumple virus*-Nirulas [India:New Delhi:Papaya:2007]; LK028571, *Hollyhock yellow vein mosaic Islamabad virus*, isolate 3AK; LM645009, *Hollyhock yellow vein mosaic Islamabad virus*, clone 2YI; KR052159, *Papaya leaf crumple virus* isolate Mohali, clone Pap-Moh7; KX302711, *Papaya leaf crumple virus* isolate Kolkata clone B2_4N; LM645010, *Cestrum yellow leaf curling virus*, clone CN-146; DQ116884, *Tomato leaf curl Pakistan virus* isolate Rahim Yar Khan 1 clone PT7; FM164938, *Tomato leaf curl Pakistan virus*, clone MI38; KP698312. *Papaya leaf crumple virus* isolate C1; LN864816, *Papaya leaf curl Faisalabad virus*, isolate ZF-19; LK028574, *Papaya leaf curl Faisalabad virus*, isolate Pakistan, Faisalabad; KX168427, *Pedilanthus leaf curl virus*; KY978406, *Pedilanthus leaf curl virus* isolate 248ISBD; KT159766, *Euphorbia yellow leaf curl virus* isolate PK1A; JN807764, *Pedilanthus leaf curl virus*-[Crape Jasmine]; LN717037, *Cestrum leaf curl Islamabad virus*, isolate CS-152; KM978186, *Euphorbia yellow leaf curl virus* isolate PK1; FM877858, *Chili leaf curl India virus*; KT948069, *Duranta leaf curl virus* isolate 57SA; AM948961, *Tomato leaf curl Pakistan virus*; KP780424, *Euphorbia yellow leaf curl virus* isolate PK2; HF568781, *Pedilanthus leaf curl virus*, isolate Pakistan, Multan, clone SPA1; LN713272, *Pedilanthus leaf curl virus*, clone SAZ59; KC852148, *Euphorbia leaf curl virus* isolate Shandong; AJ558121, *Euphorbia leaf curl virus*-[G35], isolate G35; GU732204, *Tomato leaf curl Pakistan virus* isolate ToLCPKV [IN:Bih]; KM359408, *Papaya leaf crumple virus* isolate A-87; HE580234, *Catharanthus yellow mosaic virus*, clone KN4; LK028570, *Catharanthus yellow mosaic virus*, isolate DR-151

ornamental plant *Clerodendrum cyrtophyllum* was established with a Geminivirus infection from Fujian (China) with symptoms of yellow mosaic patterns. Extracted DNA-A was mainly clade with *Clerodendrum golden mosaic virus* in the phylogenetic analyses having 78.9% nucleotide sequence identity. Accordingly DNA-B was also identified suggesting being the first report of a bipartite Begomovirus infecting *Clerodendrum cyrtophyllum* in China (Yang et al. 2009). A new incidence of *Hibiscus rosa-sinensis*, an ornamental plant (Pekamwar et al. 2013), was showing yellow and leaf curl symptoms in Guangzhou, Guangdong Province of China. Comparison showed that DNA-A had more than 89% sequence identify with all isolates of *Cotton leaf curl Multan virus* (CLCuMV). Satellite DNA molecule was also found to be associated with *Hibiscus rosa-sinensis*. Pairwise comparison indicated that DNAbeta had the highest sequence identities with CLCuMV DNAbeta (92.1%). It is concluded that the Begomovirus infecting *Hibiscus rosa-sinensis* is an isolate of CLCuMV (Mao et al. 2008).

In Pakistan *Vinca minor* L., an ornamental plant (Farahanikia et al. 2011) was monitored with geminiviral symptoms near the School of Biological Sciences, Lahore, Pakistan. Genome was characterized, and computational analysis like basic local alignment search tool analysis highlights 93% nucleotide sequence identity with *Pedilanthus leaf curl virus* originating from Pakistan (Haider et al. 2008). From Pakistan, another ornamental plant, *Duranta erecta*, also known as golden dewdrop, is cultivated as hedge in the gardens (Bruggeman 1964) belonging to the family Verbenaceae, indicated reduced leaf size, leaf curling, and chlorosis. The Geminivirus sequences isolate showed similarity with *Croton yellow vein virus* and *Papaya leaf curl virus*. These results proved that the Begomovirus residing in golden dewdrop is either a recombinant virus or a distinct Begomovirus. *Duranta leaf curl virus* was the name suggested for this Geminivirus (Iram et al. 2005). In a similar report from Pakistan, *Duranta repens* (pigeon berry) (Keong 1982) from the *Verbenaceae* family infected with Begomovirus showed the highest levels of DNA-A sequence identity to *Croton yellow vein mosaic virus* (91%), whereas DNA-B showed the highest level of identity with *Tomato leaf curl New Delhi virus* segment B (94%), suggesting a bipartite Geminivirus associated with the leaf curl disease of *D. repens* (Tahir et al. 2006). In Pakistan an ornamental shrub redbird flower, also known as *Pedilanthus tithymaloides* (Srivastava et al. 1985), observed with symptoms of leaf curl and enation. Complete Begomovirus and betasatellite genome was sequenced, revealing nucleotide sequence identity to a *Pedilanthus leaf curl virus* and Tobacco leaf curl betasatellite, respectively. This Geminivirus is one of an increasing number of monopartite Begomoviruses shown to be associated with a betasatellite (Tahir et al. 2009). In another incidence, an ornamental plant *Catharanthus roseus* (Nejat et al. 2015) known as Madagascar periwinkle, well known for its medicinal uses such as anti-cancerous properties of its alkaloids, was found infected with Geminiviruses in Pakistan. Complete genome of two Begomoviruses (clone KN4 and KN6) was determined. Clone KN4 showed nearly 87% DNA sequence identity with *Chili leaf curl India virus* (ChiLCIV), and ~85% identity to *Papaya leaf curl virus* (PaLCV) represents a new species, named as *Catharanthus yellow mosaic virus* (CYMV). The sequence of another Begomovirus

(clone KN6) showed the highest level of sequence identity (~99–99%) to *Papaya leaf crumple virus* (PaLCrV) reported from India. Sequence analysis shows that KN4 and KN6 are recombinants of *Pedilanthus leaf curl virus* (PedLCV) and *Croton yellow vein mosaic virus* (CrYVMV) (Ilyas et al. 2013).

For the first time in Malaysia, a whitefly-transmitted Begomovirus was detected using serological and molecular techniques in *Mimosa invisa* (Srivastava et al. 2012). The observed symptoms in infected plants were yellowing and stunted growth. Sequence comparison showed that this Geminivirus has the greatest sequence similarity (92%) to *Ageratum yellow vein China virus* (Koravieh et al. 2008). Ha et al. (2008) identified two Begomovirus that infect ornamental plants in Vietnam. Sequence analysis of the complete genomes showed that they belong to novel species and was identified in Vietnam for the first time. This includes *Clerodendrum golden mosaic virus* (ClGMV), infecting glory bower (*Clerodendrum philippinum*) (Venkatanarasimman et al. 2012), and *Mimosa yellow leaf curl virus* (MiYLCV), infecting mimosa (*Mimosa sp.*). The DNA-1 molecule from mimosa formed a distinct branch between DNA-1 sequences originating from China. In July 2011, symptoms of mild yellowing and leaf curling were presented on an ornamental plant *Zinnia elegans* (Riaz et al. 2008) from Vietnam (Hue City). *Zinnia leaf curl virus*, *Ageratum enation virus*, *Tobacco leaf curl virus*, and *Alternanthera yellow vein virus* were found to be associated with leaf curl disease of *Zinnia elegans*. Therefore, this virus may pose a serious threat to many plants (Li et al. 2013). In an another report, complete nucleotide sequences of the Geminivirus DNA-A genome and its DNA β satellite of *Honeysuckle yellow vein mosaic virus* and *Eupatorium yellow vein virus* were identified in *Honeysuckle sp.* (Kitamura et al. 2004; Kemp et al. 1991) and ornamental plant *Eupatorium sp.* (Deepa and Jofeena 2015), respectively, in Japan. Genome pairwise comparison of DNA-A and its DNA β satellite showed that this Geminivirus had significance threshold of 84% identity. Furthermore, recombination analysis highlighted that members of this group had the genetic variation indicative of many recombination events (Ueda et al. 2008).

4.1.3 Africa

Eragrostis curvula is known as South African wild grass and also cultivated as an ornamental plant (Leigh 1961) that was found to be infected with a new highly divergent Geminivirus species *Eragrostis curvula streak virus*. *Eragrostis curvula streak virus* has features similar to other specific Geminivirus genera such as replication-associated protein (Rep) similar to those of Begomoviruses, Curtoviruses, Topocuviruses, and Mastreviruses. Likewise *Eragrostis curvula streak virus* has the same unusual TAAGATTCC virion strand replication origin nonanucleotide similar to *Beet curly top Iran virus*. The transcription and replication origin was found structurally more similar to those of Begomoviruses and Curtoviruses. *Eragrostis curvula streak virus* also had homologue of the

Begomovirus transcription activator protein gene found in Begomoviruses, a Mastrevirus-like coat protein gene and two intergenic regions (Varsani et al. 2009a). *Panicum streak virus* a Mastrevirus represents a close relative of *Maize streak virus*. Both are transmitted by the same leafhopper vector species. Sixteen new *Panicum streak virus* full-genome sequences were determined across Africa. Analysis of both the Geminiviruses suggested that the progress of *Maize streak virus* strains in whole Africa is apparently less constrained than that of *Panicum streak virus* strains. It was concluded that increased mobility of *Maize streak virus* to other closely related species such as *Panicum streak virus* may have been an important evolutionary step in the eventual emergence of *Maize streak virus* as a serious agricultural pathogen (Varsani et al. 2009b). Maize streak virus of the genus Mastrevirus is responsible for maize streak diseases of *Zea mays* across Africa. Nigeria is one of the West African countries, where maize is not cultivated throughout the year, hence Maize streak virus resides in alternative hosts like ornamental crops and weed plants. Maize streak virus isolates were obtained from maize and *Digitaria ciliaris* grass (Cardina et al. 2011). Further a novel highly divergent mastrevirus from *Axonopus compressus* (Kamal-Uddin et al. 2009) was characterized. *Axonopus compressus* is a species of grass, used a ground cover in gardens, and was named as *Axonopus compressus streak virus* (ACSV) (Oluwafemi et al. 2014).

4.1.4 Europe

Ipomoea indica, also known as blue morning glory, belongs to *Convolvulaceae* family of plants which prevails along the coast of Greek. *Ipomoea indica* grow like weeds and also as cultivated ornamentals (Palanisamy and Arumugam 2014). During the summer of 2013, yellowing of leaf veins symptoms were observed in *Ipomoea indica*, collected from Mandriko, Kremasti, and Kolymbari regions of Greece. The Geminivirus isolated from Kolymbari were ~99% similar among themselves and showed ~98% DNA identity with a *Sweet potato leaf curl virus* isolate obtained from *Ipomoea indica* plant in Sicily (Italy). The Geminivirus reported from Kremasti revealed ~93% nucleotide similarity with *Sweet potato leaf curl virus* obtained from *Ipomoea indica* plant from southern Spain. All the identified Geminiviruses infecting *Ipomoea indica* in Greece are isolates of *Sweet potato leaf curl virus* (Fiallo-Olivé et al. 2013a). Using advance and molecular techniques, two new strains of *Abutilon mosaic virus* have been characterized from Paris (France) and Stuttgart (Germany). Their phylogenetic results suggested clustering with other *Abutilon mosaic virus* strains reported from Middle American, whereas Geminiviruses from South American infecting Abutilon or *Sida micrantha* (Gomaa et al. 2016) are less closely related. By comparing the coat protein (CP) genes of the *Abutilon mosaic virus* forms a clade with Middle and South American Begomoviruses suggesting non-synonymous nucleotide exchanges for certain amino acid positions in the *Abutilon mosaic virus* cluster (Fischer et al. 2014).

Ipomoea indica, a vegetatively propagated ornamental plant, was found exhibiting yellow vein disease symptoms in Norwich, Spain. A Geminivirus (AJ132548) was identified using molecular techniques, which showed low sequence similarity with *Ageratum yellow vein virus* (AYVV), *Tomato leaf curl virus* from southern India, and other Begomoviruses. DNA-B was absent when characterized with degenerate DNA-B primers. This virus was named as *Ipomoea yellow vein virus* (IYVV). Whiteflies transmission studies were carried out on healthy *I. indica* plants using biotypes S, Q (from Spain), and B (from Israel) that showed non-transmissible nature of *Ipomoea yellow vein virus* to healthy *Ipomoea indica*, *Lycopersicum esculentum*, *Nicotiana tabacum*, or nightshade. The loss of vector transmissibility of this Geminivirus was due to many years of vegetative propagation of the host plant as an ornamental, as occurred in *Honeysuckle yellow vein mosaic virus* (HYVMV) and *Abutilon mosaic virus* (AbMV) (Banks et al. 1999). Sweet potato (*Ipomoea batatas*) is cultivated in a number of places globally and houses Geminiviruses in this vegetatively propagated cash crop (Milind and Monika 2015) causing a hinder to its production around 90% yield loss. A survey was conducted in Spain to identify Geminiviruses infecting sweet potato and *Ipomoea indica*. A total of 15 isolates were sequenced, and results showed the presence of *Sweet potato leaf curl virus* (SPLCV), *Ipomoea yellowing vein virus*, three new species of Begomovirus, and a novel strain of SPLCV. The analysis confirmed that a number of recombination events have transformed *Ipomoea* infecting Begomoviruses in Spain (Lozano et al. 2009).

4.1.5 Australia

Mastreviruses infects a wide range of monocotyledonous plants in the Old World. Mastreviruses showed maximum diversity in African subcontinent. A research was carried out in Australia for the better understanding of mastrevirus diversity in Australia. For the same 41 Mastreviruses were found naturally in native grasses in Australia. Out of them four new Mastrevirus were found in addition to the four previously characterized species. Two highly divergent Mastrevirus recovered from a single Sporobolus plant (used in garden) were most closely related to *African streak viruses* showed inter- and intraspecies recombination (Kraberger et al. 2012).

4.1.6 New World Echelon of Geminiviruses Infecting Ornamental Plants

The ornamentals are unquestionably among the crops whose gardening consequences into generation of jobs and cohort of income. Geminiviruses are known to cause massive crop yield losses across the globe (Rojas et al. 2005; Rybicki 1994). The harm caused by Geminiviruses affects both the quality and quantity of the

preferred parts of the crops. Moreover, they were found to be widely distributed in surveyed locations (Seal 2006; Stanley et al. 2005; Hanley-Bowdoin et al. 2000; Padidam et al. 1999). As the battle to eradicate Geminivirus diseases continues, the virologists and plant breeders are hereby urged not to neglect the damages on ornamentals that are caused by the viruses reported below from the New World.

4.1.7 North America

With the help of molecular techniques like rolling circle amplification, two viral genomes were amplified from salvia (*Salvia splendens*) cv. "Dancing Flame" (Błażewicz-Woźniak et al. 2012) exhibited variegated foliage resembling Geminivirus symptoms in the United States. Bioinformatic analysis suggested the positiveness of Begomovirus *Clerodendron golden mosaic China virus*. The virus was detected in all symptomatic salvia plants but absent in non-variegated plants. Further study was done on graft transmission of virus to healthy *Salvia splendens* which reproduced the original symptoms. Even attempts were made to transmit the virus with the help of whitefly, *Bemisia tabaci* biotype B, but were negative (Valverde et al. 2012). For the first time a Topocuvirus, i.e., *Tomato pseudo curly top virus* (TPCTV), has been recorded in Florida whose genome was cloned and sequenced. Topocuvirus is the only Geminivirus transmitted by treehopper vector. Infectivity assay was carried out via Agrobacterium-mediated inoculation of several host species. Descendants of the Topocuvirus were transmissible by the treehopper vector in *Micrutalis malleifera* (Fowler). The genome and coat protein of *Tomato pseudo curly top virus* show features distinct from all previously characterized Geminiviruses, revealing close relationship to *Beet curly top virus* (Briddon et al. 1996). In Ohio State of the United States, an ornamental plant "Gold Veined Oxalis" (*Oxalis debilis*) (Junejo et al. 2016) displays eye-catching yellow vein foliage which was used for graft transmission to other *Oxalis debilis* plants. DNA was isolated and subjected to polymerase chain reaction; amplified sequence revealed that a Geminivirus was responsible for yellow vein symptoms with *Bemisia tabaci* the carrier vector. The Begomovirus was suggested the name *Oxalis yellow vein virus* (OxYVV) which showed 87%, 85%, and 83% DNA sequence identities with *Sida mottle virus*, *Okra mottle virus*, and with *Tomato yellow spot virus*, respectively, confirming another instance of a Geminivirus that augments the aesthetic beauty of ornamental plants (Herrera et al. 2015).

Mexico has been a house of number of local and cultivated eudicots (Worberg et al. 2007) in its Yucatan Peninsula and was found with Geminivirus symptoms. Thirteen different Geminiviruses were identified mainly Begomoviruses. Out of all, five were new Begomovirus species named as *Boerhavia yellow spot virus* (BoYSV), *Anoda golden mosaic virus* (AnGMV), *Desmodium leaf distortion virus* (DeLDV), *Hibiscus variegation virus* (HiVV), and *Papaya golden mosaic virus* (PaGMV). Already identified species of Begomoviruses were *Sida golden mosaic virus* (SiGMV), *Melon chlorotic leaf curl virus* (MCLCuV), *Euphorbia mosaic virus*

(EuMV), *Tobacco apical stunt virus* (TbASV), and *Okra yellow mosaic Mexico virus* (OkYMMV) which were found for the first time. Finally viruses earlier reported in Yucatan Peninsula were *Tomato mottle virus* (ToMoV), *Pepper golden mosaic virus* (PepGMV), and *Bean golden yellow mosaic virus* (BGYMV) which were recognized in eudicots plants. Their phylogenetic analysis grouped all Begomoviruses in Western Hemisphere clade (Hernández-Zepeda et al. 2007).

4.1.8 South America

The major destructive group of viruses infecting cultivated crops across the globe is Geminiviruses. In a study in Boqueirão which is place in Paraíba state of Brazil, leaf samples were sampled from *Sida* spp. (Gomaa et al. 2016) and tomato crops. Geminiviruses were isolated, and 30 DNA-A were identified. Computational analysis reveals only two Begomovirus, *Sida mottle Alagoas virus* (SiMoAV) infecting *Sida* spp. and *Tomato mottle leaf curl virus* (ToMoLCV) obtained from tomato crops. Rep gene of ToMoLCV showed the highest variability and recombination events in comparison to SiMoAV. This study highlights ToMoLCV as the chief Begomovirus harboring tomatoes in northeastern side of Brazil (Ferro et al. 2016). Moreover there are also reports of Geminiviruses that are responsible for variegation in flower that comprises beautiful mosaic patterns caused by *Abutilon mosaic virus* mainly on the leaves of flowering maple (*Abutilon pictum*; Malvaceae) (Święczkowska and Kowalkowska 2015). This plant is also known as parlor maple or Indian mallow and used as an ornamental plant in America. "Gold Dust" and "Thompsonii" are the cultivar used for parlor maple. Initially parlor maple belongs to Brazil and now prevails across the South and Central America and has been growing for ornamental purpose since the 1800s (Keur 1934). Even Morren in 1869 was successfully transmitted the variegation caused by Geminiviruses through graft inoculations. Even in recent years, there have been other reports of plant viruses associated with unique phenotypes of ornamental plants, and some of them are being marketed as novel plant cultivars (Valverde et al. 2011). Whiteflies transmitted Geminiviruses are responsible for serious infections in crops and are frequently also coupled with non-cultivated plants. Two new Geminiviruses were identified in *Pavonia* sp. (da Silva et al. 2016) belonging to Malvaceae family in Brazil. When their sequence comparisons were carried out and further phylogenetic analysis was done, it was revealed that these novel Geminiviruses correspond to the New World Begomoviruses. The DNA sequences of the DNA-A of the two viruses revealed the highest similarity to *Abutilon mosaic Bolivia virus* (AbMBoV). Based on the symptoms identified in the field and taking into account the host plant, the names *Pavonia mosaic virus* (PavMV) and *Pavonia yellow mosaic virus* (PavYMV) were proposed for these two new Geminiviruses (Pinto et al. 2015).

Geminiviruses plays a crucial role in the crop production worldwide; hence it's a necessity to detect the viral pathogen at an early stage of plant infection. *Abutilon mosaic virus* (AbMV) is a Begomovirus species and belongs to *Geminiviridae*

family of plant viruses, which known to infect ornamental plants across the globe. In the garden area of Tumbaco (Equador) *Abutilon hybridum* plants (Holser and Bost 2004) were observed with mosaic patterns bright yellow in color as a sign of infection symptoms. The Geminivirus was identified (KP877621) by the help of degenerate primers against Begomovirus in polymerase chain reaction showing 68 to cent percent nucleotide identity with other Begomoviruses. A novel method using Raman spectroscopy was performed acting as a diagnostic tool for AbMV infection by recording the spectrum from leaves of both healthy and symptomatic samples. In leaves carotenoids form the basis for Geminivirus identification by showing a sharp decline in the intensity of the bands. The change in the intensity of the bands, especially at 1526 cm^{-1}, was designated as the basis for earlier detection of Geminiviral infection in plants (Yeturu et al. 2016). Even the complete sequences of new bipartite Geminivirus infecting *Datura stramonium* in Venezuela were also obtained. The name proposed for this new Begomovirus were *Datura leaf distortion virus* (DLDV) which were obtained from *Datura stramonium* L. plant belonging to *Solanaceae* family (Hansen and Clerc 2002). The dendrogram revealed that *Datura leaf distortion virus* separates out in two divergent clades of New World Begomoviruses (Khurana and Marwal 2016; Fiallo-Olivé et al. 2013b).

4.1.9 Management Strategies Against Geminiviruses

Conventional strategies can fail to control rapidly evolving and emerging plant viruses. Genome engineering strategies have recently emerged as promising tools to introduce desirable traits in many eukaryotic species, including plants (Zaidi et al. 2016). Transgene-induced silencing in plants is usually associated with methylation of nuclear DNA corresponding to the transcribed region of the target RNA despite transcription levels of the transgene remain unaffected (Saunders et al. 2004). RNAi technology when used against a Geminivirus (*African cassava mosaic virus* (ACMV)) showed 99% decrease of Rep transcripts and 66% reduction in viral DNA (Sanjaya et al. 2005). Targeting Rep and AV_2 gene by antisense technology is found to be quite successful (Dasgupta et al. 2003). To make PTGS an effective method, both sense and antisense RNAs are a prerequisite, and a transgenic tobacco (*Nicotiana benthamiana*) using RNAi was developed (Singh et al. 2007). Sense and antisense constructs that were made of the movement protein genes (BC1 and BV1) of tomato mottle Geminivirus (TMoV) were transformed by Agrobacterium-mediated transformation in tobacco explants. Geminivirus was detected in inoculated leaves but was not readily detected in leaves beyond the inoculation sites in the highly resistant plants (Duan et al. 1997). A mastrevirus *Maize streak Reunion virus* (MSRV) collected in Yunnan Province, China, was identified by small RNA deep sequencing. This vsiRNAs profile derived from MSRV-YN was characterized, which might contribute to get an insight into the host RNA-silencing defense induced by MSRV-YN and provide guidelines on designing antiviral strategies using RNAi against MSRV-YN (Marwal and Gaur 2017; Chen et al. 2015).

Level of disease control obtained using traditional measures led to the search for other control measures that go beyond traditional host genetic resistance, chemical controls, and cultural practices. Recently, the CRISPR/Cas technology has emerged as potent genome editing tool, with remarkable applications in various organisms, including several plant species. The CRISPR/Cas system originates from prokaryotic organisms and acts as an adaptive immune system to protect them against invading foreign DNA, such as phages, by cleaving the nucleic acid by an RNA-guided DNA nuclease in a sequence-specific manner (Chaparro-Garcia et al. 2015). A virus-based guide RNA (gRNA) delivery system for CRISPR/Cas9-mediated plant genome editing (VIGE) that causes mutations in target genome locations has been reported. VIGE was executed by using a modified *Cabbage leaf curl virus* (CaLCuV) vector to express gRNAs in stable transgenic plants expressing Cas9 (Yin et al. 2015). CRISPR–Cas is an adaptive immune system in many archaea and bacteria that cleave foreign DNA on the basis of sequence complementarity. Geminiviruses causes great crop losses worldwide. In order to tackle such problem, a study was carried out by Baltes et al. (2015) in bean yellow dwarf virus (BeYDV) genome which was targeted for destruction with the CRISPR–Cas system. By using BeYDV-based replicons, transient assays were done disclosing that CRISPR–Cas reagents introduced mutations within the viral genome and thus reduced virus copy number. Transgenic plants infected with BeYDV showed less virus load and symptoms by expressing CRISPR–Cas reagents. This novel strategy demonstrates a perfect engineering resistance to Geminiviruses (Baltes et al. 2015). Similarly, using the Geminivirus, *Beet severe curly top virus* (BSCTV), transient assays performed in *Nicotiana benthamiana* which demonstrate that the sgRNA–Cas9 constructs inhibit virus accumulation and introduce mutations at the target sequences. Further, transgenic *Arabidopsis* and *N. benthamiana* plants overexpressing sgRNA–Cas9 are highly resistant to virus infection (Ji et al. 2015).

4.2 Conclusion

Throughout life, we are encircled by an indistinguishable stratum of atmosphere. We know that the atmosphere contains copious microscopic particles such as pollens, allergens, dust particles, and species of microorganisms ranging from bacteria to fungi and viruses. And the viruses are considered the foremost environmental danger to humans, animals, and plants. Several viruses affect the crop plants, but the major contribution is from the viruses belonging to the *Geminiviridae* family (Gaur et al. 2011). In recent years the Geminiviruses has received a great deal of attention, and it is one of the most important studied plant viruses. Several reviews on Geminiviruses covered different aspects of their biology, epidemiology, serological properties, and molecular biology (Prajapat et al. 2014). Tropical and subtropical regions are the most favorable regions for emerging of newly Geminiviruses which cause severe disease epidemics. Ornamentals can retain these Geminiviruses that can be transmitted by the insect vector back to crop plants causing yield loss of the crops. Once

present in the new host, these viruses would have rapidly evolved giving rise to new devastating species. Ornamental plants forming one of the major branches of horticulture plays a huge role by supplementing the income of the rural people. The production of these plants becomes key drivers of economic development in many parts of the world. These ornamental plants serve as an alternative host for the viruses in gardens, and they may facilitate Geminivirus transmission to crop plants, thus enhancing the host range of this virus in different regions of the world (Marwal et al. 2016). Current trends for extensive and intensive agriculture, open international agricultural trade, food security, and sustained economy have brought new challenges in the fight against Geminivirus diseases. For this reason, this chapter focused on and presents the scenario of various reports of characterization and identification of different Geminiviruses infecting in ornamental plants worldwide.

Acknowledgments The authors are thankful to Science and Engineering Research Board – Department of Science and Technology, New Delhi, India, for the financial assistance (File No. YSS/2015/000265) and also to the University Grant Commission, New Delhi, for providing financial assistantship under Research Award for Teacher (F.30-1/2014/RA-2014-16-GE-RAJ-4696 (SA-II).

References

Ali A, Ahmed M, Nishigawa H, Natsuaki T (2014) Identification of *Tobacco leaf curl virus* infecting *Lonicera japonica*, an ornamental plant common in Japan. J Agric Sci Technol 16:645–655

Banks GK, Bedford ID, Beitia FJ, Rodriguez-Cerezo E, Markham PG (1999) A novel geminivirus of *Ipomoea indica* (Convolvulacae) from Southern Spain. Plant Dis 83(5):486

Baldotto LEB, Baldotto MA (2015) Growth and production of ornamental sunflower grown in the field in response to application of humic acids. Ciênc Rural Santa Maria 45(5):1000–1005

Baliji S, Black MC, French R, Stenger DC, Sunter G (2004) *Spinach curly top virus*: a newly described Curtovirus species from Southwest Texas with incongruent gene phylogenies with incongruent gene phylogenies. Phytopathology 94:772–779

Briddon RW, Bedford ID, Tsai JH, Markham PG (1996) Analysis of the nucleotide sequence of the treehopper-transmitted geminivirus, *Tomato Pseudo Curly Top Virus*, suggests a recombinant origin. Virology 219:387–394

Bruggeman L (1964) Tropical plants and their cultivation. Thames and Hudson, London, p 228

Baltes NJ, Hummel AW, Konecna E, Cegan R, Bruns AN, Bisaro DM, Voytas DF (2015) Nat Plants. https://doi.org/10.1038/nplants.2015.145

Błażewicz-Woźniak M, Madej J, Rtemi D, Wartacz W (2012) The growth and flowering of *Salvia splendens* Sellow ex Roem. et Schult under flowerbed conditions. Acta Agrobot 65(2):99–108

Cardina J, Herms CP, Herms DA (2011) Phenological indicators for emergence of large and smooth crabgrass (*Digitaria sanguinalis* and *D. ischaemum*). Weed Technol 25(1):141–150. 2011

Chethana GS, Hari Venkatesh KR, Gopinath SM (2013) Review on *Clerodendrum inerme*. J Pharm Sci Innov JPSI 2(2):38–40

Chaparro-Garcia A, Kamoun S, Nekrasov V (2015) Boosting plant immunity with CRISPR/Cas. Genome Biol 16:254

Chaveerach A, Aungkapattamagu S, Tanee T, Noikotr K, Sudmoon R (2014) Genetic verification and chemical contents identification of Allamanda species (*Apocynaceae*). Pak J Pharm Sci 27 (3):417–424

Chen S, Huang Q, Wu L, Qian Y (2015) Identification and characterization of a maize-associated mastrevirus in China by deep sequencing small RNA populations. Virol J 12:156

Dasgupta I, Malathi VG, Mukherjee SK (2003) Genetic engineering for virus resistance. Curr Sci 84:341–354

Deepa GM, Jofeena J (2015) Analysis of phytochemicals from *Eupatorium odoratum* flower. Int J Ayurveda Pharm Res 3(1):1–5

Duan YP, Powel CA, Webb SE, Purcifull DE, Hiebert E (1997) Geminivirus resistance in transgenic tobacco expressing mutated BC1 protein. MPMI 10(5):617–623

Farahanikia B, Akbarzadeh T, Jahangirzadeh A, Yassa N, Ardekani MRS, Mirnezami T, Hadjiakhoondi A, Khanavi M (2011) Phytochemical investigation of *Vinca minor* cultivated in Iran. Iran J Pharm Res 10(4):777–785

Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA (2011) Family *Geminiviridae*. In: Virus taxonomy: classification and nomenclature of viruses: Ninth report of the international committee on taxonomy of viruses. Elsevier, San Diego, pp 301–326

Ferro MMM, Ramos-Sobrinho R, Silva JT, Assunção IP, Lima GSA (2016) Genetic structure of populations of the Begomoviruses *Tomato mottle leaf curl virus* and *Sida mottle Alagoas* virus infecting tomato (*Solanum lycopersicum*) and Sida spp., respectively. Trop Plant Pathol. https://doi.org/10.1007/s40858-016-0119-z

Fiallo-Olivé E, Chirinos DT, Geraud-Pouey F, Moriones E, Navas-Castillo J (2013a) Complete genome sequences of two Begomoviruses infecting weeds in Venezuela. Arch Virol 158:277–280

Fiallo-Olivé E, Katis NI, Navas-Castillo J (2013b) First report of *Sweet potato leaf curl virus* on blue morning glory in Greece. Plant Dis 98(5):700

Fischer A, Strohmeier S, Krenz B, Jeske H (2014) Evolutionary liberties of the *Abutilon mosaic virus* cluster. Virus Genes. https://doi.org/10.1007/s11262-014-1125-1

Garcia-Arenal F, Fraile A, Malpica JM (2001) Variability and genetic structure of plant virus populations. Annu Rev Phytopathol 39:157–186

Gaur RK, Prajapat R, Marwal A, Sahu A, Rathore MS (2011) First report of a begomovirus infecting *Mimosa pudica* in India. J Plant Pathol 93(SUPPL. 4):S4.63–S4.89

Gilbertson RL, Rojas MR, Russell DR, Maxwell DP (1991) Use of asymmetric polymerase chain reaction and DNA sequencing to determine genetic variability of *bean golden mosaic* geminivirus in the Dominican Republic. J Gen Virol 72:2843–2848

Gilbertson RL, Hidayat SH, Paplomatas EJ, Rojas MR, Hou YH, Maxwel DP (1993) Pseudorecombination between infectious cloned DNA components of tomato mottle and bean dwarf mosaic Geminiviruses. J Gen Virol 74:23–31

Gomaa AA, Samy MN, Desoukey SY, Kamel MS (2016) Pharmacognostical studies of leaf, stem, root and flower of *Abutilon hirtum* (Lam.) sweet. Int J Pharm Phytochem Res 8(1):199–216

Govindappa MR, Shankergoud I, Shankarappa KS, Wickramaarachchi WART, Reddy BA, Rangaswamy KT (2011) Molecular detection and partial characterization of Begomovirus associated with leaf curl disease of sunflower (*Helianthus annuus*) in Southern India. Plant Pathol J 10:29–35

Ha C, Coombs S, Revill P, Harding R, Vu M, Dale J (2008) Molecular characterization of Begomoviruses and DNA satellites from Vietnam: additional evidence that the New World Geminiviruses were present in the Old World prior to continental separation. J Gen Virol 89:312–326

Haider MS, Tahir M, Saeed A, Ahmed S, Parveen R, Rashid N (2008) First report of a Begomovirus infecting the ornamental plant *Vinca minor* L. Aust Plant Dis Notes 3:150–151

Hamdy RS, Abd El-Ghani MM, Youssef TL, El-Sayed M (2007) The floristic composition of some historical botanical gardens in the metropolitan of Cairo, Egypt. Afr J Agric Res 2(11):610–648

Hansen P, Clerc B (2002) Anisocoria in the dog provoked by a toxic contact with an ornamental plant: *Datura stramonium*. Vet Ophthalmol 5(4):277–279

Hanley-Bowdoin L, Settlage SB, Orozco BM, Nagar S, Robertson D (1999) Geminiviruses: models for plant DNA replication, transcription and cell cycle regulation. Crit Rev Plant Sci 18:71–106

Hanley-Bowdoin L, Settlage SB, Orozco BM, Nagar S, Robertson D (2000) Geminiviruses: models for plant DNA replication, transcription, and cell cycle regulation. Crit Rev Biochem Mol Biol 35:105–140

Hanley-Bowdoin L, Bejarano ER, Robertson D, Mansoor S (2013) Geminiviruses: masters at redirecting and reprogramming plant processes. Nat Rev Microbiol. https://doi.org/10.1038/nrmicro3117

He ZF, Mao MJ, Yu H, Li HP, Chen X (2009) Molecular characterization of a distinct Begomovirus infecting *Allamanda cathartica* in Guangdong, China. Arch Virol 54:1199–1202

Hernández-Zepeda C, Idris AM, Carnevali G, Brown JK, Moreno-Valenzuela OA (2007) Preliminary identification and coat protein gene phylogenetic relationships of Begomoviruses associated with native flora and cultivated plants from the Yucatan Peninsula of Mexico. Virus Genes 35:825–833

Hernandez-Zepeda C, Varsani A, Brown JK (2013) Intergeneric recombination between a new, spinach-infecting Curtovirus and a new Geminivirus belonging to the genus Becurtovirus : first New World exemplar. Arch Virol. https://doi.org/10.1007/s00705-013-1733-4

Herrera F, Aboughanem-Sabanadzovic N, Valverde RA (2015) A Begomovirus associated with yellow vein symptoms of *Oxalis debilis*. Eur J Plant Pathol. https://doi.org/10.1007/s10658-015-0594-y

Herrera-Vásquez JA, Córdoba-Sellés MC, Cebrián MC, Font-San-Ambrosio MI, Alfaro-Fernández A, Jordá C (2013) Viruses of cucurbits in Panama. J Plant Pathol 95:435–440

Heydarnejad E, Abhari H, Yazdi HRB, Massumi H (2007) Curly top of cultivated plants and weeds and report of a unique curtovirus from Iran. J Phytopathol 155:321–325

Holser RA, Bost G (2004) Hybrid hibiscus seed oil compositions. JOACS 8:795–797

Ilyas M, Nawaz K, Shafiq M, Haider MS, Shahid AA (2013) Complete nucleotide sequences of two Begomoviruses infecting Madagascar periwinkle (*Catharanthus roseus*) from Pakistan. Arch Virol 158:505–510

Iram S, Amrao L, Mansoor MS, Malik AH, Briddon RW, Zafar Y (2005) First report of a Begomovirus associated with leaf curl disease of *Duranta erecta* in Pakistan. Plant Pathol 54:260

Ji X, Zhang H, Zhang Y, Wang Y, Gao C (2015) Establishing a CRISPR–Cas-like immune system conferring DNA virus resistance in plants. Nat Plants. https://doi.org/10.1038/nplants.2015.144

John P, Sivalingam PN, Kumar N, Mishra A, Ahlawat YS, Malathi VG (2006) A new Begomovirus associated with yellow mosaic disease of *Clerodendron inerme*. Plant Pathol 55:291

Junejo JA, Gogoi G, Zaman K, Islam J, Hazarika G (2016) Phytochemical profiling and toxicological studies of *Oxalis debilis* Kunth leaves. Int J Green Pharm 10(3):165–171

Kamal-Uddin M, Juraimi AS, Begum M, Ismail MR, Rahim AA, Othman R (2009) Floristic composition of weed community in turf grass area of west peninsular Malaysia. Int J Agric Biol 11:13–20

Kareem A, Saeed S, Hammad HM (2014) Growth and performance of *Calendula officinalis* L. on different crop residues. World J Agric Sci 2(5):098–101

Kemp WP, Berry JS, Caprio JM (1991) Use of ornamental lilac and honeysuckle phenophases as indicators of rangeland grasshopper development. J Range Manag 44(6):583–586

Keong WC (1982) Rooting cuttings of *Gardenia jasminoides*, *Duranta repens*, and *Bougainville glabra* with growth retardant. PeItanika 5(1):25–29

Keur JY (1934) Studies of the occurrence and transmission of virus diseases in the genus Abutilon. Bull Torrey Bot Club 61:53–70

Khan AA, Naqvi QA, Khan MS, Singh R, Raj SK (2005) First report of a Begomovirus infecting Calendula in India. New Dis Rep 54:569

Khurana SMP, Marwal A (2016) Recent developments towards detection & diagnosis for management of plant viruses. Indian Phytopathol 69(4s):30–34

Kitamura K, Murayama A, Ikegami M (2004) Evidence of recombination among isolates of *Tobacco leaf curl Japan virus* and *Honeysuckle yellow vein mosaic virus*. Arch Virol 149:1221–1229

Konate G, Barro N, Fargette D, Swanson MM, Harrison BD (1995) Occurrence of whitefly-transmitted Geminiviruses in crops in Burkina Faso, and their serological detection and differentiation. Ann Appl Bid 126:121–129

Koravieh MM, Roff MNM, Othman RY (2008) First report of a whitefly-transmitted Geminivirus infecting *Mimosa invisa* in Malaysia. Aust Plant Dis Notes 3:25–26

Kraberger S, Thomas JE, Geering ADW, Dayaram A, Stainton D, Hadfield J, Walters M, Parmenter KS, van Brunschot S, Collingsa DA, Martin DP, Varsani A (2012) Australian monocot-infecting mastrevirus diversity rivals that in Africa. Virus Res 169:127–136

Lapido JI (1988) Viruses of vegetable crops in Africa. In: Williams AO, Mbiele AL, Nkouka N (eds) Virus diseases of plants in Africa. OAU/STRC Scientific Publication, Lagos, pp 157–167

Leigh JH (1961) The relative palatability of various varieties of weeping love grass (*Eragrostis curvula* (Schrad.) Nees). J Br Grassl Soc 16:135–140

Li J, Zhou X (2010) Molecular characterization and experimental host-range of two Begomoviruses infecting *Clerodendrum cyrtophyllum* in China. Virus Genes 41:250–259

Li ZB, Qin BX, Cai JH (2013) First report of *Ageratum yellow vein China virus* infecting *Zinnia elegans* in Vietnam. New Dis Rep 97:431

Lozano G, Trenado HP, Valverde RA, Navas-Castillo J (2009) Novel Begomovirus species of recombinant nature in sweet potato (*Ipomoea batatas*) and *Ipomoea indica*: taxonomic and phylogenetic implications. J Gen Virol 90:2550–2562

Mahesh YS, Shankarappa KS, Rangaswamy KT, Prameela HA, Aswathanarayana DS, Maruthi MN (2008) Detection of a begomovirus associated with leaf curl disease of ornamental croton (*Codieum variegatum*) in southern India. J Plant Pathol 90:143–149

Mahesh YS, Shankarappa KS, Rangaswamy KT, Prameela HA, Aswathanarayana DS, Divya BL, Nagesha N, Maruthi MN (2010) Detection and characterisation of a Begomovirus associated with leaf curl disease of ornamental croton (*Codiaeum variegatum*). J Hortic Sci Biotechnol 85 (2):101–105

Mahesh YS, Shankarappa KS, Rangaswamy KT, Prameela HA, Aswathanarayana DS, Divya BL, Nagesha N, Maruthi MN (2012) Detection and characterisation of a Begomovirus associated with leaf curl disease of ornamental croton (*Codiaeum variegatum*). J Hortic Sci Biotechnol 85:101–105

Mahgoub MH, Nahed, Aziz GA, Abu-Zied KM, Mohamed TK (2015) Effect of khellinone derivatives on growth and flowering of *Dianthus chinensis* L. Plant Int J Adv Res. (2015) 3 (10):130–138

Mao MJ, He ZF, Yu H, Li HP (2008) Molecular characterization of *Cotton leaf Curl Multan virus* and its satellite DNA that infects *Hibiscus rosa-sinensis*. Bing Du Xue Bao 24:64–68

Marwal A, Gaur RK (2017) Understanding functional genomics of PTGS silencing mechanisms for Tobacco streak virus and other Ilarviruses mediated by RNAi and VIGS. Plant-Microbe Interactions in Agro-Ecological Perspectives. In: Dhananjaya Pratap Singh (ed) Volume 1: Fundamental mechanisms, methods and functions. Springer, Chapter 24. pp 489–499, DOI: https://doi.org/10.1007/978-981-10-5813-4_24

Marwal A, Sahu A, Prajapat R, Choudhary DK, Gaur RK (2012) First report of association of Begomovirus with the leaf curl disease of a common weed, *Datura inoxia*. Virus Dis 23 (1):83–84

Marwal A, Sahu A, Choudhary DK, Gaur RK (2013a) Complete nucleotide sequence of a Begomovirus associated with satellites molecules infecting a new host *Tagetes patula* in India. Virus Genes 47(1):194–198

Marwal A, Sahu A, Sharma P, Gaur RK (2013b) Molecular characterizations of two Begomoviruses infecting *Vinca rosea* and *Raphanus sativus* in India. Virol Sin 28(1):053–056

Marwal A, Sahu A, Sharma P, Gaur RK (2013c) Transmission and host interaction of geminivirus in weeds. In: Plant virus-host interaction: molecular approaches and viral evolution. Elsevier Chapter 7, pp 143–161

Marwal A, Sahu A, Gaur RK (2014) First report of airborne Begomovirus infection in *Melia azedarach* (Pride of India), an ornamental tree in India. Aerobiologia 30(2):211–215

Marwal A, Prajapat R, Gaur RK (2016) First report of recombination analysis of betasatellite and alphasatellite sequence isolated from an ornamental plant marigold in India: an in silico approach. Int J Virol 12(1–3):10

Marwal A, Mishra M, Sekhsaria C, Gaur RK (2017) Computational analysis and predicting ligand binding site in the Rose leaf curl virus and its betasatellite proteins: A step forward for antiviral agent designing. In: Saxena S, Tiwari AK (eds) Begomoviruses: Occurrence and management in Asia and Africa, 1st edn. Springer, Chapter 9, pp 157–168. https://doi.org/10.1007/978-981-10-5984-1_9

Milind P, Monika (2015) Sweet potato as a super food. Int J Res Ayurveda Pharm 6(4):557–562

Moffat AS (1999) Plant pathology – geminiviruses emerge as serious crop threat. Science 286:1835

Morales FJ, Anderson PK (2001) The emergence and dissemination of whitefly-transmitted geminiviruses in Latin America. Arch Virol 146:415–441

Morren E (1869) Contagion de la panachure. Bull Acad R Sci 28:434–442

Muhire B, Martin DP, Brown JK, Navas-Castillo J, Moriones E, Zerbini FM, Rivera-Bustamante R, Malathi VG, Briddon RW, Varsani A (2013) A genome-wide pairwise-identity-based proposal for the classification of viruses in the genus mastrevirus (*Geminiviridae*). Arch Virol. https://doi.org/10.1007/s00705-012-1601-7

Navot N, Pichersky R, Zeidan M, Zamir D, Czosnek H (1991) *Tomato yellow leaf curl virus*: a whitefly-transmitted geminivirus with a single genomic component. Virology 185:151–161

Nejat N, Valdiani A, Cahill D, Tan YH, Maziah M, Abiri R (2015) Ornamental exterior versus therapeutic interior of Madagascar periwinkle (*Catharanthus roseus*): the two faces of a versatile herb. Sci World J. https://doi.org/10.1155/2015/982412

Oluwafemi S, Kraberger S, Shepherd DN, Martin DP, Varsani A (2014) A high degree of *African streak virus* diversity within Nigerian maize fields includes a new mastrevirus from *Axonopus compressus*. Arch Virol. https://doi.org/10.1007/s00705-014-2090-7

Openshaw K (2000) A review of *Jatropha curcas*: an oil plant of unfulfilled promise. Biomass Bioenergy 19:1–15

Osaki T, Inouye T (1983) Whitefly transmitted viruses isolated form sweet potato, *Abutilon pictum* and *Eupatorium japonica*. 1. Resemblance in intranuclear appearance of viruses. Ann Phytopathol Soc Jpn 52:128

Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. Virology 265:218–225

Palanisamy J, Arumugam R (2014) Exploration of wild ornamental flora of Madukkarai hills of southern western Ghats, Tamil Nadu. Biolife 2(3):834–841

Pandey N, Tiwari AK (2012) Identification of *Zinnia leaf curl virus* infecting *Zinnia elegans* in India. J Biotechnol Bioinform 2:6–10

Pekamwar SS, Kalyankar TM, Jadhav AC (2013) *Hibiscus rosa-sinensis*: a review on ornamental plant. World J Pharm Pharm Sci 2(6):4719–4727

Polston JE, Anderson PK (1997) The emergence of whitefly-transmitted geminiviruses in tomato in the western hemisphere. Plant Dis 81:1358–1369

Prajapat R, Marwal A, Sahu A, Gaur RK (2011) Phylogenetics and in silico docking studies between coat protein of *Mimosa yellow vein virus* and whey α-lactalbumin. Am J Biochem Mol Biol 1(3):265

Prajapat R, Marwal A, Shaikh Z, Gaur RK (2012a) Geminivirus database (GVDB): first database of family *Geminiviridae* and its genera Begomovirus. Pak J Biol Sci 15(14):702

Prajapat R, Marwal A, Sahu AK, Gaur RK (2012b) Molecular in silico structure and recombination analysis of betasatellite in *Calotropis procera* associated with Begomovirus. Arch Phytopathol Plant Protect 45(16):1980

Prajapat R, Marwal A, Gaur RK (2014) Begomovirus associated with alternative host weeds: a critical appraisal. Arch Phytopathol Plant Protect 47(2):157

Pruss G, Ge X, Shi MX, Carrington JC, Vance VB (1997) Plant viral synergism: the potyviral genome encodes a broad-range pathogenicity enhancer that transactivates replication of heterologous viruses. Plant Cell 9:859–868

Pinto VB, Silva JP, Fiallo-Olive E, Navas-Castillo J, Zerbini FM (2015) Novel Begomoviruses recovered from *Pavonia sp*. in Brazil. Arch Virol. https://doi.org/10.1007/s00705-015-2708-4

Raj SK, Khan MS, Snehi SK, Kumar S, Khan AA (2007) Natural occurrence of a Begomovirus on *Dimorphotheca sinuate* in India. Aust Plant Dis Notes 2:25–26

Riaz A, Arshad M, Younis A, Raza A, Hameed M (2008) Effects of different growing media on growth and flowering of *Zinnia elegans*. Pak J Bot 40(4):1579–1585

Rojas MR, Hagen C, Lucas WJ, Gilbertson RL (2005) Exploiting chinks in the plant's armor: evolution and emergence of geminiviruses. Annu Rev Phytopathol 43:361–394

Rybicki EP (1994) A phylogenetic and evolutionary justification for 3 genera of *Geminiviridae*. Arch Virol 139:49–77

Rybicki EP, Hughes FL (1990) Detection and typing of *Maize streak virus* and other distantly related geminiviruses of grasses by polymerase chain reaction amplification of a conserved viral sequence. J Gen Virol 71(2519):2526

Sahu AK, Marwal A, Nehra C, Shahid MS, Gaur RK (2014) First report of a begomovirus and associated betasatellite in *Rosa indica* and in India. Aust Plant Dis Notes 9:147

Sanjaya VVS, Prasad V, Kirthi N, Maiya SP, Savithri HS, Sita GL (2005) Development of cotton transgenics with antisense AV2 gene for resistance against *Cotton leaf curl virus* (CLCuD) via *Agrobacterium tumefaciens*. Plant Cell Tissue Organ Cult 81:55–63

Saunders K, Norman A, Gucciardo S, Stanley J (2004) The DNA-b satellite component associated with ageratum yellow vein disease encodes an essential pathogenicity protein (bC1). Virology 324:37–47

Seal SE, vanden Bosch F, Jeger MJ (2006) Factors influencing Begomovirus evolution and their increasing global significance: implications for sustainable control. Crit Rev Plant Sci 25:23–46

Singh DK, Karjee S, Malik PS, Islam N, Mukherjee SK (2007) DNA replication and pathogenicity of MYMIV. In: Mendez-Vilas A (ed) Communicating current research and educational topics and trends in applied microbiology. Formatex, Spain, pp 155–162. ISBN-13: 9788461194223

da Silva TBC, Costa COD, Galvão AFC, Bomfim LM, da C. Rodrigues ACB, MCS M, Dantas AA, dos Santos TR, MBP S, Bezerra DP (2016) Cytotoxic potential of selected medicinal plants in Northeast Brazil. BMC Complement Altern Med 16:199

Snehi KS, Raj SK, Prasad V, Singh V (2016) Molecular detection and identification of Begomovirus isolates associated with mosaic disease of ornamental Jatropha species from India. Int J Bacterio Virol 1:004

Soto MJ, Chen LF, Seo YS, Gilbertson RL (2005) Identification of regions of the beet mild curly top virus (family *Geminiviridae*) capsid protein involved in systemic infection, virion formation and leafhopper transmission. Virology 341:257–270

Speight JG, Singh K (2014) Environmental management of energy from biofuels and biofeedstocks. Wiley, New York. https://doi.org/10.1002/9781118915141

Srivastava GS, Bhatiaa VK, Dubey KC, Garg VK (1985) Potential of *Pedilanthus tithymaloides* as a petro-crop. Fuel 64(5):720–721

Srivastava V, Sharma A, Alam I (2012) A review on ethnomedical and traditional uses of *Mimosa pudica* (chui-mui). Int Res J Pharm 3(2):41

Srivastava A, Snehi SK, Raj SK (2013) First report of a variant of *Pepper leaf curl Lahore virus* on winter cherry in India. New Dis Rep 27:7

Stanley J, Bisaro DM, Briddon RW, Brown JK, Fauquet CM, Harrison BD, Rybicki EP, Stenger DC (2005) Family geminiviridae. In: Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA (eds) Virus taxonomy: eighth report of the international committee on taxonomy of viruses. Elsevier Academic Press, London, pp 301–326

Święczkowska E, Kowalkowska AK (2015) Floral nectary anatomy and ultrastructure in mycoheterotrophic plant, *Epipogium aphyllum* Sw. (*Orchidaceae*). Sci World J. https://doi.org/10.1155/2015/201702

Tahir M, Haider MS, Shah AH, Rashid N, Saleem F (2006) First report of bipartite Begomovirus associated with leaf curl disease of *Duranta repens* in Pakistan. New Dis Rep 88:339–342

Tahir M, Haider MS, Iqbal J, Briddon RW (2009) Association of a distinct Begomovirus and a betasatellite with leaf curl symptoms in *Pedilanthus tithymaloides*. J Phytopathol 157:188–193

Thomas JE, Massalski PR, Harrison BD (1986) Production of monoclonal antibodies to *African cassava mosaic virus* and differences in their reactivities with other whitefly-transmitted geminiviruses. J Gen Virol 67:2739–2748

Ueda S, Onuki M, Hanada K, Takanami Y (2008) Unique grouping of the far East Asian Begomovirus complex based on sequence analyses of the DNA-A genome and associated DNA-β satellite molecules isolated from tomato, honeysuckle and Eupatorium plants in Japan. Arch Virol 153:417–426

Uga TS (2005) A one-step reverse transcription-polymerase chain reaction system for the simultaneous detection and identification of multiple tospovirus infections. Phytopathology 95 (2):166–171

Valverde RA, Sabanadzovic S, Hammond J (2011) Viruses that enhance the aesthetics of some ornamental plants: beauty or beast? Plant Dis 96:600–611

Valverde RA, Singh R, Sabanadzovic S (2012) Detection and identification of *Clerodendron golden mosaic China virus* in *Salvia splendens*. Eur J Plant Pathol 133:499–503

Varsani A, Monjane AL, Donaldson L, Oluwafemi S, Zinga I, Komba EK, Plakoutene D, Mandakombo N, Mboukoulida J, Semballa S, Briddon RW, Markham PG, Lett JM, Lefeuvre P, Rybicki EP, Martin DP (2009a) Comparative analysis of *Panicum streak virus* and *Maize streak virus* diversity, recombination patterns and phylogeography. Virol J 6:194

Varsani A, Shepherd DN, Dent K, Monjane AL, Rybicki EP, Martin DP (2009b) A highly divergent South African geminivirus species illuminates the ancient evolutionary history of this family. Virol J 6:36

Varsani A, Navas-Castillo J, Moriones E, Herna´ndez-Zepeda C, Idris A, Brown JK, Zerbini FB, Martin DP (2014) Establishment of three new genera in the family *Geminiviridae*: becurtovirus, eragrovirus and turncurtovirus. Arch Virol 159:2193–2203

Venkatanarasimman B, Rajeswari T, Padmapriya B (2012) Preliminary phytochemical screening of crude leaf extract of *Clerodendrum philippinum* Schauer. Int J Inst Pharm Life Sci 2(2):133–138

Weinberger K, Msuya J (2004) Indigenous vegetables in Tanzania, significance and prospects. AVRDC The World Vegetable Centre. Tech Bull 31:71

Worberg A, Quandt D, Barniske AM, Lohne C, Hilu KW, Borsch T (2007) Phylogeny of basal eudicots: insights from non-coding and rapidly evolving DNA. Org Divers Evol 7:55–77

Yin K, Han T, Liu G, Chen T, Wang Y, Yunzi A, Yu L, Liu Y (2015) A geminivirus-based guide RNA delivery system for CRISPR/Cas9. Sci Rep 5:14926. https://doi.org/10.1038/srep14926

Yang CX, Luo JS, Hong JT, Xie LH, Wu ZJ (2009) *Clerodendrum yellow mosaic China virus* is a distinct bipartite Begomovirus isolated from *Clerodendrum Cyrtophyllum* in China. J Plant Pathol 91(4):S4.97–S4.112

Yeturu S, Jentzsch PV, Ciobota V, Guerrero R, Garrido P, Ramos LA (2016) Handheld Raman spectroscopy for the early detection of plant diseases: *Abutilon mosaic virus* infecting Abutilon sp. Anal Methods. https://doi.org/10.1039/C6AY00381H

Zaidi SS, Tashkandi M, Mansoor S, Mahfouz MM (2016) Engineering plant immunity: using CRISPR/Cas9 to generate virus resistance. Front Plant Sci 7:1673. https://doi.org/10.3389/fpls.2016.01673

Chapter 5
In Silico Methods to Predict Disease-Resistance Candidate Genes in Plants

Jyoti Lakhani, Ajay Khuteta, Anupama Choudhary, and Dharmesh Harwani

Abstract The increase in world population is stressing the need for increased production of food supplies from plants. At the same time, plant pathogens are also developing resistance to several anti-pathogen compounds. The present situation may become worse, in the near future, if not controlled. One of the solutions to the present situation is to develop disease-resistant varieties of plants. The disease resistance in plants is controlled by the products of disease-resistance genes. The plant genomes contain many disease-resistance candidate genes, activation of which can confer the natural resistance against various diseases in plants. The major step in the development of disease-resistant plant varieties is to search for the disease-resistance candidate genes in the plant genome and prioritize them. The experiments pertaining to identify disease-resistance candidate genes can be accomplished using wet lab studies but are usually time-consuming. The present chapter is a survey of the available in silico approaches to identify the candidate genes conferring disease resistance in plants. After providing a brief overview of the multilayered defense mechanism, the present article discusses different approaches for the stepwise identification of disease-resistant candidate genes in plants.

Keywords Plant genome · Disease resistance · Candidate genes · Genome mining · In silico prediction

J. Lakhani
Department of Computer Science, Maharaja Ganga Singh University, Bikaner, Rajasthan, India

Poornima University, Jaipur, India

A. Khuteta
Department of Computer Science, Poornima University, Jaipur, India

A. Choudhary
Department of Computer Science, Keen College, Bikaner, India

D. Harwani (✉)
Department of Microbiology, Maharaja Ganga Singh University, Bikaner, Rajasthan, India
e-mail: dharmesh@mgsubikaner.ac.in

D. K. Choudhary et al. (eds.), *In Silico Approach for Sustainable Agriculture*,
https://doi.org/10.1007/978-981-13-0347-0_5

5.1 Introduction

Production of sufficient amount of crops, proportional to feed the global population, is a big challenge. The loss of yield of approximately 20–40% has been seen because of the pathogen assaults in most important crops such as wheat, rice, and maize (Rawat 2016). Out of this, approximately 15% global loss has been observed due to various plant diseases (Patnala et al. 2013). Plants are being constantly subjected to biotic and abiotic stresses. Their response to the pathogen attack has been studied extensively and systematically (Dang et al. 2013). A three-tier immune system in plants resists the pathogens. In the first tier, plants recognize pathogen molecules outside the cell wall. This process is also called pattern-triggered immunity (PTI). In this initial defense reaction of plant immune system toward the pathogen attack, pathogen- derived molecules called pathogen-associated molecular patterns (PAMPs) or microbe-associated molecular patterns (MAMPs) are recognized. This is performed by pathogen recognition receptors (PRRs) present on the plasma membrane (Dang et al. 2013) in plants. In the second tier, plants recognize intruders inside the cells, and this specific defense is called as effector-triggered immunity (ETI). The PTI and ETI receptors initiate oxidative burst, ion channel or kinase cascades, several changes in cellular redox, and gene expression. The role of the third tier of immunity mechanism is to trigger the defense mechanism in the other infected tissue and neighboring cells by sending alert messages. In some cases, defense-activating signals spread to the rest of the plant or to the neighboring plants in the vicinity. The last tier is shared with the first and second tiers to channelize the overall defense system (Jones and Dangl 2006; Dodds and Rathjen 2010). This complete defense system in plants has been elaborated in Fig. 5.1.

The second level of plant defense mechanism is completely depended on resistance (R) genes. The plant immune systems carry a repertoire of 100–600 different R gene homologs. The plants have evolved with R genes, product(s) of which mediates resistance to the specific pathogens such as viruses, bacteria, fungi, insects, etc. The product of R genes are proteins that allow recognition of specific pathogen effectors, either through direct binding or by recognizing effector's alteration of the host protein (Jones and Dangl 2006). The system is triggered by the specific receptors encoded by pathogen recognition genes (PRGs). The proteins encoded by the PRGs share coiled-coil (CC), nucleotide-binding region (NB), toll/interleukin region (TIR), leucine-rich region (LRR), and kinase (K) domain. The cytoplasmic NB-LRR genes are divided into two classes, TNL (TIR-NB-LRR) and CNL (CC-NB-LRR), which possess either TIR or CC domains. The transmembrane receptor proteins, possessing kinase and LRR domains, such as receptor-like proteins (RLP) and receptor-like kinases (Sharma et al. 2014), are also reported to be involved (Andolfo and Ercolan 2015). The R genes also encode NB-LRR proteins with nucleotide-binding and leucine-rich repeat domains. Specialized plant pathogens can evade or suppress MAMP-triggered immunity (MTI) by secretion of the virulence factors called effectors. A subset of these effectors known as avirulence factors (AVRs) can be recognized by the resistance proteins that trigger the second

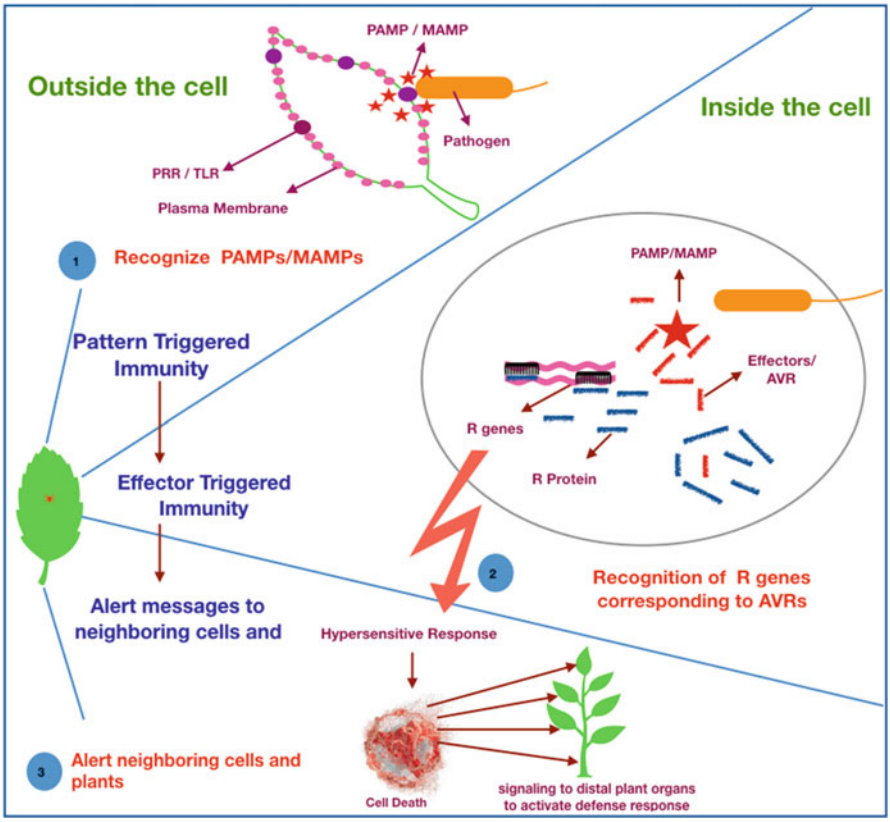


Fig. 5.1 Multilayered defense system in plants. The figure shows three tiers of plant defense system against pathogens denoted by one, two, and three. The first tier in the above figure displays pattern-triggered immunity. This level of immunity is triggered when a pathogen comes into the contact with the plant. The membrane secreted pathogen recognition receptor (PPR) recognizes pathogen-associated molecular patterns (PAMPs)/microbe-associated molecular patterns (MAMPs), and as a result, the second level of immunity is triggered. In the plant genomes, pathogen recognition genes (PRGs) detect PAMPs/MAMPs. The pathogen produces effectors/AVR, after getting the entry inside the plant cell which in turn triggers R genes to secrete R proteins. The rapid increase in this activity results in the hypersensitivity response which leads to the cell death and the activation of the third tier of immunity. Consequently, third tier of immunity-driven signals is sent to the neighboring cells or distal plant organs to activate the defense system in advance against the invading pathogen

layer of the host defense which is referred as effector-triggered immunity (ETI) or R gene-mediated defense. Consequently, this is characterized by a rapid, localized cell death at the infection site which is termed as hypersensitivity response (HR). The HR leads to the signaling to the distal plant organs to activate the defense response genes which confer the protection against subsequent infection by the pathogen (Fig. 5.1).

The pioneering work based on the genetics of plant disease resistance involving HR was reported by Harold H. Flor in linseed *Melampsora lini* (Flor 1956). Flor

proposed a gene-for-gene hypothesis that classically demonstrated the genetic interaction between a plant and a pathogen. The hypothesis states that for every resistance gene in a plant, there is a corresponding avirulence gene (*avr*) in the pathogen. The interaction between the host resistance gene and pathogen avirulence gene leads to the incompatibility (resistance). The model hypothesizes that there could be a direct or indirect physical interaction between the ligand produced by pathogen and corresponding plant receptor which ultimately triggers the activation of downstream defense response genes.

Martin et al. (1993) provided the first evidence for direct interaction of tomato *pto* gene with *avrPto* from *Pseudomonas syringae* pv. tomato. Apart from direct interaction, the report indicates that resistance proteins (RPs) also act as the guard of specific component of the basic defense pathway (guard hypothesis). If that component is modified by the pathogen effector molecule, the modification is recognized by R protein, thus activating the defense response (Van Der Biezen and Jones 1998). The disease-resistance genes have been extensively reviewed elsewhere (Bent 1996; Van Der Biezen and Jones 1998; Martin 1999; Michelmore 2000; Ellis et al. 2000; Jones 2001; Meyers et al. 2005) underpinning its importance in sustainable agriculture. It is clear that R genes impart an important role in the plant defense mechanism against pathogens.

5.2 Prediction of Disease-Resistance Candidate Genes

A disease-resistance candidate gene (DRCG) is a gene, believed to confer resistance to a particular disease or condition. DRCGs can be identified on the basis of its position in the plant genome or by its functional analysis (Fig. 5.2). A positional DRCG is located in the chromosomal region which is supposed to be linked to the disease resistance. A functional DRCG is a gene, encoding a protein directly linked to the disease resistance (Hale et al. 2005). A variety of search methods are in place for in silico prediction of DRCGs for their known position in the genome, or otherwise, as discussed above, they can be identified also by their gene expression analysis.

5.3 Selection of Putative DRCGs

The proper selection of a suitable disease-resistance gene is a fundamental step in the prediction process. The DRGCs' selection can be carried out in the following ways.

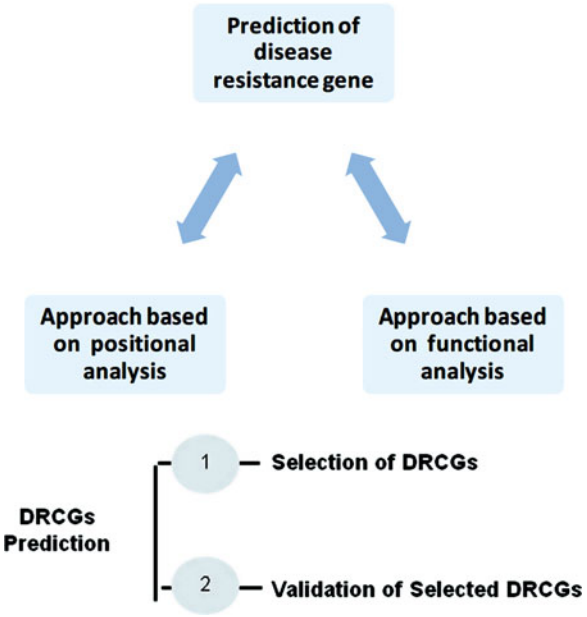


Fig. 5.2 Approaches for DRCG prediction, selection, and validation

5.3.1 Literature Mining

This method is based on the conventional literature mining from the most common biological databases such as PubMed (https://www.ncbi.nlm. nih.gov/pubmed/), BioText (Hearst et al. 2007), Google Scholar (https://scholar.google.co.in) etc. There are some other text mining tools available on the web that assist in literature mining. The iHOP web service (Fernandez et al. 2012) tool assists in general literature mining; PubCrawler (Hokamp and Wolfe 2004) helps to keep the track of daily updates. Biovista (https://www.biovista.com) is helpful in visualizing the results of the search made by the user. The traditional literature mining method is suitable for the well-studied biological problems concerned with single model species (Slota et al. 2017). These conventional methods are not suitable for complex problems, where the multiple traits or under-examined traits are considered. These complex problems can be targeted by performing the empirical or in silico analyses. The other in silico text mining tools are PLAN2L (Krallinger et al. 2009), Beegle (Sarah et al. 2015), pGenN (Ding et al. 2015), Dragon Plant Biology Explorer (Bajic et al. 2005), ePlant (Jamie Waese et al. 2017), etc.

5.3.2 *In Silico Resources*

The disease-resistance gene (DRG) can be searched from the plant resistance gene (PRG) databases which have been specifically designed and maintained for the data pertaining to the plant R genes (Sanseverino et al. 2010a, b). These are manually curated databases that hold well-characterized plant disease-resistance gene belonging to nearly 200 plant species. The available in silico resources can be categorized into four classes (Slota et al. 2017). These resources are used in the positional and functional analysis of DRG. The web-based tools are helpful in searching literature-driven data from the web-based repositories. General annotations and characterization of the biological and molecular functions of DRCGs have been already reported using web-based tools by Slota group (2017). Some genomic repositories can also be used for retrieving homologous genes from the different plant species. There are several other online gene expression data repositories that contain hybridization, microarrays, and RNA-sequence data. The databases such as Gene Expression Omnibus (GEO) (Edgar and Domrachev 2002), ArrayExpress (Parkinson et al. 2007; Brazma et al. 2003), or Genevestigator (Zimmermann et al. 2004) are some other examples of publicly available, gene expression data repositories. Gene expression databases lead to the identification of differentially expressed gene subsets across the different experimental conditions. The identification of gene expression profiles can significantly reinforce the process for the selection of DRCGs associated with a distinct biological process or environmental stimuli (Slota et al. 2017). DRCGs can be searched according to their functional annotation, molecular characteristics, or protein localization from the gene ontology (GO) databases (Harris et al. 2004). GO database is a unified vocabulary system and relies on the characterization of biological properties and hierarchical classification of gene products in the three clusters: cellular component, molecular function, and biological process involvement (Carbon et al. 2009a, b). Plant gene ontology database can be accessed from agriGO (Du et al. 2010), AmiGO Gene Ontology (Carbon et al. 2009b), Plant Ontology (Cooper and Jaiswal 2016), and QuickGO (Binns et al. 2009; Huntley et al. 2009) browsers.

5.3.3 *Gene Prediction Algorithms (QTL-Based Approach)*

The gene prediction algorithms are used in position-based identification methods. FGENESH (Salamov and Solovyev 2000) is such HMM-based gene structure prediction algorithm (multiple genes, both chains). Gene finder is considered as the most accurate tool for the identification of plant genes. Chaudhari and Fakrudin (2016) have predicted DRCGs in the quantitative trait loci (QTL) regions of the mapped genome of sorghum, and the observed range of length of the QTL intervals

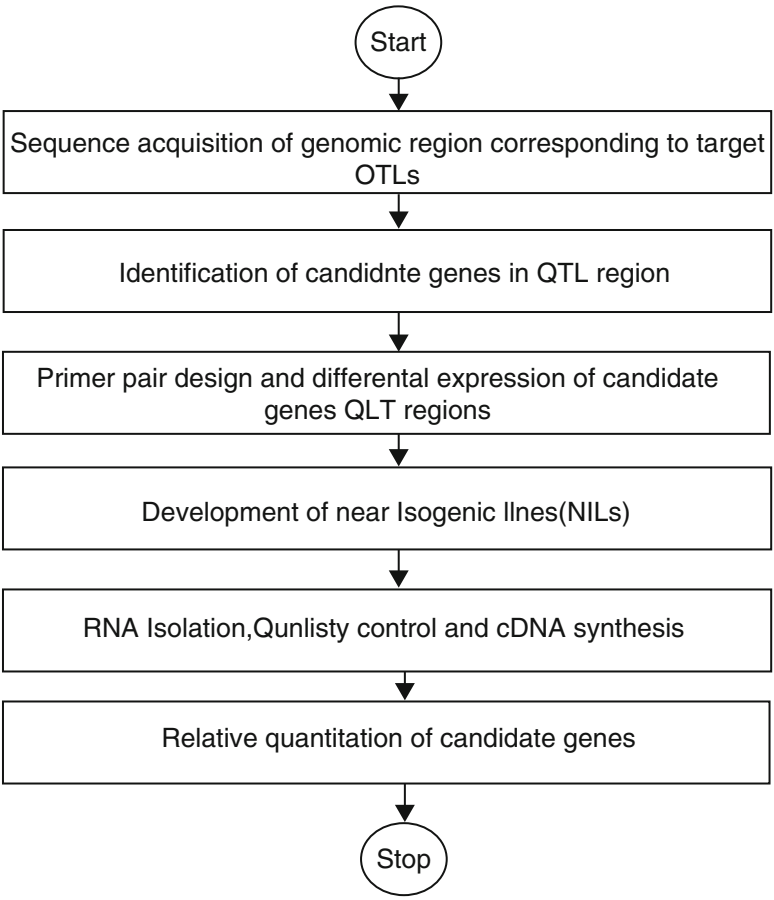


Fig. 5.3 The process of identification of DRCGs in QTL region

was from 1.8 to 2.5 Mbp. The other gene prediction algorithms are GENSCAN (Burge and Karlin 1997) and GenMark (Besemer and Borodovsky 2005). These two algorithms cumulatively predicted a total of *nsp*, *nad*, *phd*, *mads*, *mlo* resistant genes in QTL qSTG1A (1.82 Mbp), *iaa*, *sorbi-draft*, *cyp450*, *gag/pol*, *pk*, *gene x*, *ugts*, *mtc*, *agp16*, *vp25* genes in qSTG2 (2.54 Mbp), and *sfcc1* gene in qSTG3 (2.18 Mbp) on chromosomes 3 and 1. These algorithms have also been used in the gene prediction in rice, barley, maize, *A. thaliana*, etc. (Hong et al. 2005; Champoux et al. 1995). Zhang group (2008) used FGENESH version 2.0 to predict genes in rice plant. The predicted genes were mostly multi-exon genes. The approach has been demonstrated in Fig. 5.3.

5.3.4 Programming Scripts (PRG-Based Approach)

Andolfo group (2013) identified the resistance gene candidates in tomato (*Solanum lycopersicum* L.) plant. For the physical mapping of the predicted pathogen recognition genes, SQL database was developed in which the genes are catalogued on the basis of their characteristics and location. Perl (Wall 1988) script has been also developed for converting sequences of each chromosome in vector graphics (SVG). To use this approach in a convincing way, one should have the detailed knowledge of programming, and therefore, the approach is not considered much useful for biologists.

5.3.5 Resistance Gene Analog (RGA-Based Approach)

Resistance gene analog (RGA) approach which is known also as in silico pipeline approach performs both selection and validation of DRCGs on a single platform. RGAugury pipeline approach (Li et al. 2016) is based on RGA recognition. RGAs such as NBS-encoding proteins, receptor-like protein kinases (RLKs), and receptor-like proteins (RLPs) are potential R genes that contain specific conserved domains and motifs. The approach resembles the functional analysis approach of DRCG identification which was developed to automate RGA prediction. The pipeline approach first identifies RGA-related protein domains and motifs such as nucleotide-binding site (NB-ARC), leucine-rich repeat (LRR), transmembrane (TM), serine/threonine and tyrosine kinase (STTK), lysin motif (LysM), coiled-coil (CC), and toll/interleukin-1 receptor (TIR) (Li et al. 2016). Thereafter, RGA candidates are identified and classified into four major families based on the presence of combinations of RGA domains and motifs as NBS-encoding, TM-CC, and membrane-associated RLP and RLK. The approach automates and parallelizes all the time-consuming steps to improve performance. The pipeline approach has been evaluated using the well-annotated *Arabidopsis* genome. A user-friendly web interface has been implemented to ease the command-line operations, facilitate visualization, and simplify the result management for multiple datasets. RGAugury is an efficient integrative bioinformatics tool for the large-scale genome-wide identification of RGAs. It is freely available at Bitbucket: https://bitbucket.org/yaanlpc/rgaugury (Li et al. 2016). ASPic-GeneID (Alioto et al. 2013) is a lightweight pipeline for gene prediction and isoform detection. This pipeline has two components. The first is ASPic, a program that derives highly accurate, albeit not necessarily complete, EST-based transcript annotations from EST alignments. The second component is GeneID which is a modified version of gene prediction program proposed by Parra et al. (2000). This modified program can also function in the annotations of introns.

5.3.6 Gene Prediction Programs

Several useful gene prediction programs have been reported in the literature. GeneID (Parra et al. 2000) is a programming tool to predict genes, exons, splice sites, and other signals in DNA sequences. JIGSAW (Allen et al. 2005) is a program which uses a statistical algorithm that predicts gene models using the output from other annotation software. AUGUSTUS (Stanke and Morgenstern 2005; Stanke et al. 2006) is an open-source program that predicts genes in the eukaryotic genome. It has been tested in *Arabidopsis thaliana*, *Chlamydomonas reinhardtii*, *Galdieria sulphuraria*, *Nicotiana tabacum*, *Solanum lycopersicum*, and *Zea mays*. EuGene (Schiex et al. 2001; Foissac et al. 2008) is an open integrated, gene finder tool for eukaryotic and prokaryotic genomes. It has the ability to integrate information from several sources for prediction process.

5.4 Validation of R Genes

The validation of predicted genes can be performed using several methods, some of which are discussed in the following section.

5.4.1 Candidate Gene Association Search

The information related to gene associations is available in experimental data repositories of different species. The gene association repositories contain a variety of databases including gene-phenotype links, gene-disease associations, gene expression and co-expression, allelic information, effects of genetic variation, links to scientific literature, homology information from model species, protein-protein interactions, gene regulation, protein pathway memberships, gene ontology annotations, protein-domain information, etc. The available experimental data can be helpful to select the resistance gene candidates by searching gene associations (Patnala et al. 2013). For this purpose, a systematic prioritization is required for successful analysis, but it is difficult to mine multiple databases at the same time to analyze these associations.

5.4.2 DRCG Expression Analysis

Ingenuity Knowledge Base (https://ingenuity.net.au/support/knowledge-base/) is a depot of manually sorted, enriched, and reviewed information of the biological interactions and functional annotations. Another website, ToppGene Suite (Chen

et al. 2009), provides tools for the functional enrichment of genes based on the training gene set (to be provided by the user) and also includes analysis of the protein networks and locus-specific neighboring genes. Prior knowledge of gene function and structural elements within its periphery can elicit a better understanding of the putative function of the gene variants. The comprehensive websites such as Entrez Gene at http://www.ncbi.nlm.nih.gov/Entrez (Martin 1999) and Ensembl at http://www.ensembl.org (Michelmore 2000) provide general information on gene structure, expression, splice variant-encoded proteins, and regulatory elements. There are many established reports which are based on the validation of the identified DRCGs using expression analysis. Alfando group (2013) performed the molecular validation of the predicted resistance gene candidates using expression data of *S. lycopersicum* obtained from the tomato genome consortium (The Tomato Genome Consortium 2012).

5.4.3 Prioritization Tools

A systematic prioritization of the candidate genes is required to plot a hypothesis to link the genotype to pathogen-derived molecules. There are several web-based resources for prioritizing candidate genes by exploiting multiple data sources (Moreau and Tranchevent 2012; Bornigen et al. 2012). The biological knowledge used for the prioritization of the disease-resistance candidate genes may include some indirect relationship parameters which need integration of the data from various sources. This is not considered as a trivial process as the data being integrated can be in different formats, and thus the integration should be avoided. However, this plays an important role in gene prioritization and hypothesis generation. Most of the gene prioritization tools such as Endeavour (Tranchevent et al. 2016) and BioGraph (Liekens et al. 2011) are restricted to analyst and key model species, but intriguingly, the data integration process is not applicable to other species under observation. The first tool to prioritize candidate genes for two different plants is PosMed-Plus (Makita et al. 2009). The approach was developed for *Arabidopsis thaliana* and rice species which is also able to provide co-occurrence and cross-species information. KnetMiner (Hassani-Pak and Rawlings 2017; Hassani 2016) tool is used for the discovery and evaluation of candidate genes from large-scale integrations of the model and non-model species from public and private databases. The Endeavour (Tranchevent et al. 2016) integrates a total of 75 datasets from 6 model species including human and mouse into a local database. The machine learning techniques are used with a priori candidate gene datasets for prioritization. BioGraph (Liekens et al. 2011) is a data warehouse approach which uses unsupervised data mining technique. The gene prioritization portal at http://www.esat.kuleuven.be/gpp such as GeneRank (Morrison et al. 2005), GeneWanderer (Sebastian et al. 2008), Caesar (Daniela Nitsch et al. 2011), SNPs3D (Yue et al. 2006), and GeneDistiller (Seelow et al. 2008), among many others, link-out different 33 computational tools for this purpose.

5.4.4 GO Database

The tools which rely on gene ontology (GO) to characterize the gene information are limited because the GO annotation is an ongoing process and does not provide complete details. In addition to this, it is biased towards well-characterized diseases and research terms hence fail to spot the specific search, which could otherwise be functionally relevant to the genes under investigation (Zhang et al. 2012). A very good care should be taken to use the most updated versions of the tools which are available online, as these have been fine-tuned to provide better accuracy rates. These tools are also supported by the latest database resources such as 1000 Genome Project (with relation to genomic variants) and genome assembly (GRCh37/hg19). Genome reference consortium website can be tracked for the latest updates at http://www.ncbi.nlm. nih.gov/projects/genome/assembly/grc/.

5.4.5 Phylogenetic Analysis

The evolutionary analyses of the predicted pathogen recognition proteins such as NBS, eLRR-Ser/Thr, and KIN groups based on the WAG model (Whelan and Goldman 2001) has been performed by Andolfo group using the maximum likelihood method in MEGA5 (Tamura et al. 2011) tool. The current version of MEGA tool is MEGA6.0 (Tamura et al. 2013). The evolutionary history has been inferred by developing the bootstrap consensus tree (Felsenstein 1985) of 100 replicates. All the amino acid sequences were aligned using MUSCLE 3.6 (Edgar 2004). Predictive network integrates gene interaction and networks information from PubMed literature and other online biological databases.

5.4.6 Pathway Analysis

Reactome (Croft et al. 2011) is a manually curated, cross-referenced, and peer-reviewed pathway database. LitInspector (Frisch et al. 2009) and NetPath (Kandasamy et al. 2010) allow access to the curated signal transduction related literature and interaction pathways. GeneGo at http://www.genego.com/ facilitates pathway analysis to find the interacting molecules and subsequent interactions relevant to the trait or disease under investigation. Ingenuity IPA tool also considers pathway analysis. Some other pathway analysis tools are GeneGo/MetaCore (www.genego.com), Ingenuity Pathway Analysis (www.ingenuity.com), Pathway Studio (www. ariadnegenomics.com), GenMAPP (www. genmapp.com), WikiPathways (www. wikipathways.org), cPath (cbio.mskcc.org/cpath), BioCyc (www.biocyc.org), Pubgene (www.pubgene.org), PANTHER (www. pantherdb.org), WebGestalt (bioinfo.vanderbilt.edu/webgestalt/), ToppGene Suite (/toppgene.cchmc.org/),

DAVID (david.abcc.ncifcrf.gov/), and Pathway Painter (pathway.painter.gsa-online.de/). In addition, some useful protein-protein interaction databases, BIND, MINT, HPRD, MPact, DIP, IntAct, PDZBase, GNPV, BioGRID, UniHi, and OPHID; metabolic pathways databases, EcoCyc, MetaCyc, and BioCyc; and signaling pathways databases, KEGG, PANTHER, Reactome, BioModels, STKE, PID, BioPP, etc., are also available online for pathway analysis.

5.5 Conclusion

Disease-resistance genes in plants function at the second tier of the multilayer immune system. R genes are responsible for the detection of pathogen molecules or effecters inside the cell. R genes encode R proteins that can detect effectors and suppress their activities which results in hypersensitivity response, leading to the cell death. The signals are then communicated to the neighboring cells or other organs to activate the defense response in advance to resist the pathogen. In this way, the disease-resistant genes play an important role in the plant protection. It is imperative to identify the appropriate R gene candidates, deployment of which enhances the active-immunity of plants. The present communication highlighted some of the most common in silico prediction techniques which are used to identify disease-resistance candidate genes (DRCGs). The method involves two-step process in which the resistant genes are identified first and then verified using one of the in silico methods described above. In silico methods for the prediction of disease-resistance candidate genes (Fig. 5.4) are more reliable and accurate than the wet lab experimental analysis. It is important to note that experimental analysis in the laboratory is not

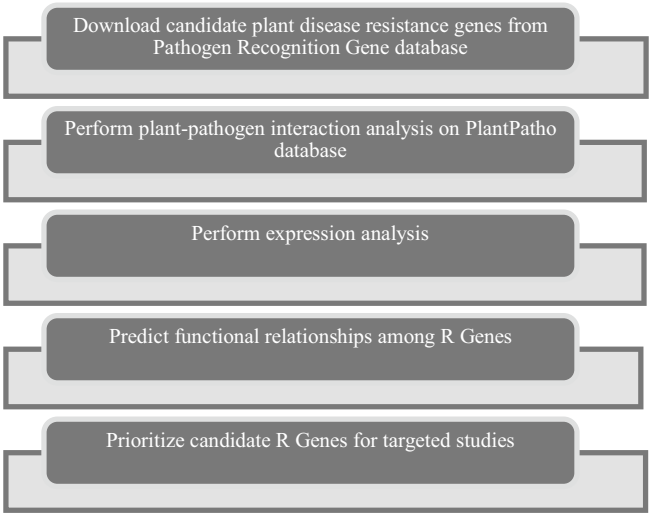


Fig. 5.4 Suggested steps for disease-resistance candidate genes

only time-consuming, but the analyses based on structure and function of genes is also tedious. In comparison to that, in silico approaches including gene prediction pipeline are user-friendly using which both the identification and validation of disease-resistance candidate genes can be performed on a single platform.

References

Alioto T, Picardi E, Guigó R, Pesole G (2013) ASPic-GeneID: a lightweight pipeline for gene prediction and alternative isoforms detection. Biomed Res Int 2013:502827. https://doi.org/10.1155/2013/502827

Andolfo G, Ercolan MR (2015) Plant innate immunity multicomponent model. Front Plant Sci 6:987–994

Andolfo G, Sanseverino W, Aversano R, Frusciante L, Ercolano MR (2013) Genome-wide identification and analysis of candidate genes for disease resistance in tomato. Mol Breed. https://doi.org/10.1007/s11032-013-9928-7

Bajic VB, Veronika M, Veladandi PS, Meka A, Heng MW, Rajaraman K, Pan H, Swarup S (2005) Dragon plant biology explorer: a text-mining tool for integrating associations between genetic and biochemical entities with genome annotation and biochemical terms lists. Plant Physiol. https://doi.org/10.1104/pp.105.060863

Bent AF (1996) Plant disease resistance genes: function meets structure. Plant Cell 8:1757–1771

Besemer J, Borodovsky M (2005) GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res 33:W451–W454. https://doi.org/10.1093/nar/gki487. http://exn.gatech.edu/eukhmm.cgiBiovista: https://www.biovista.com

Binns D et al (2009) QuickGO: a web-based tool for gene ontology searching. Bioinformatics 22:3045–3046

Bornigen D, Tranchevent L-C, Bonachela-Capdevila F, Devriendt K, De Moor B, De Causmaecker P et al (2012) An unbiased evaluation of gene prioritization tools. Bioinformatics 28:3081–3088

Brazma A et al (2003) ArrayExpress—a public repository for microarray gene expression data at the EBI. Nucleic Acids Res 31(1):68–71

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. J Mol Biol 268(1):78–94. https://doi.org/10.1006/jmbi.1997.0951. http://genes.mit.edu/GENSCAN.html

Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S (2009a) AmiGO: online access to ontology and annotation data. Bioinformatics 25(2):288–289

Carbon S et al (2009b) AmiGO: online access to ontology and annotation data. Bioinformatics 25 (2):288–289

Champoux MC, Wang G, Sarkarung S, Mackill DJ, O'Toole JC, Huang N, McCouch SR (1995) Locating genes associated with root morphology and drought avoidance in rice via linkage to molecular markers. Theor App Genet 90:969–981

Chaudhari GN, Fakrudin B (2016) Candidate gene prediction and expression profiling of near isogenic lines (NILs) carrying stay-green QTLs in rabi sorghum. J Plant Biochem Biotechnol. https://doi.org/10.1007/s13562-016-0362-x

Chen J, Bardes EE, Aronow BJ, Jegga AG (2009) ToppGene suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res 37(suppl 2):W305–W311

Cooper L, Jaiswal P (2016) The plant ontology: a tool for plant genomics. In: Edwards D (ed) Plant bioinformatics, Methods in Molecular Biology, vol 1374. Humana Press, New York

Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, Caudy M, Garapati P, Gopinath G, Jassal B et al (2011) Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res 39(suppl 1):D691–D697

Dang JL, Horvath DM, Stakawicz BJ (2013) Pivoting the plant immune system from dissection to deployment. Science 341:746–751

Ding R, Arighi CN, Lee JY, Wu CH, Vijay-Shanker K (2015) pGenN, a gene normalization tool for plant genes and proteins in scientific literature. PlosOne. https://doi.org/10.1371/journal.pone.0135305

Dodds PN, Rathjen JP (2010) Plant immunity: towards an integrated view of plant–pathogen interactions. Nat Rev Genet 11(8):539–548. https://doi.org/10.1038/nrg2812.PMID20585331

Du Z et al (2010) agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res 38(Web Server issue):W64–W70

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5):1792–1797

Edgar R, Domrachev LAE (2002) Gene expression omnibus: NCBI gene expression and hybridization Array data repository. Nucleic Acids Res 30(1):207–210

Ellis J, Dodds P, Pryor T (2000) Structure, function and evolution of plant disease resistance gene. Curr Opin Plant Biol 3:278–284

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x

Fernández J, Hoffmann R, Valencia A (2012) In: Freitas A, Navarro A (eds) iHOP web services family in bioinformatics for personalized medicine, vol 6620. Springer, Berlin, pp 102–107

Flor (1956) The complementary genic systems in flax and flax rust. Adv Genet 8:29–54

Foissac S, Gouzy J, Rombauts S, Mathé C, Amselem J, Sterck L, Peer YV, Rouzé P, Schiex T (2008) Genome annotation in plants and Fungi: EuGène as a model platform. Curr Bioinforma 3:87–97

Frisch M, Klocke B, Haltmeier M, Frech K (2009) LitInspector: literature and signal transduction pathway mining in PubMed abstracts. Nucleic Acids Res 37(suppl 2):W135–W140. Google Scholar: https://scholar.google.co.in

Hale WG, Saunders VA, Margham JP (2005) Collins dictionary of biology, 3rd edn. London, Collins

Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, et al., Gene ontology consortium (2004) The gene ontology (GO) database and informatics resource. Nucleic Acids Res 32:D258–D261

Hassani K (2016) Developing integrated crop knowledge networks to advance candidate gene discovery. Appl Transl Genomics 11:18–26. https://doi.org/10.1016/j.atg.2016.10.003

Hassani-Pak K, Rawlings C (2017) Knowledge discovery in biological databases for revealing candidate genes linked to complex phenotypes. J Integr Bioinform 14:0002

Hearst MA, Divoli A, Guturu H, Ksikes A, Nakov P, Wooldridge MA, Ye J (2007) BioText search engine: beyond abstract search. Bioinformatics 23:2196–2197. https://omictools.com/biotext-search-engine-tool

Hokamp K, Wolfe KH (2004) PubCrawler: keeping up comfortably with PubMed and GenBank. Nucleic Acids Res 32(suppl 2):W16–W19

Hong Y, Ling G, Yan F, Lisa A, Tsui-Jung B, Wen Devid S, Xiangqin S, Cuil Brain E, Scheffer JC, Scott J, Emrich DA, Ashlock Patrick SS (2005) Evaluation of five ab initio gene prediction programs for the discovery of maize genes. Plant Mol Biol 57:445–460

Huntley RP, Binns D, Dimmer E, Barrell D, O'Donovan C, and Apweiler R (2009) QuickGO: a user tutorial for the web-based gene ontology browser. Database (Oxford). https://doi.org/10.1093/database/bap010

Jones JDG (2001) Putting knowledge of plant disease resistance genes to work. Curr Opin Plant Biol 4:281–287

Jones JD, Dangl JL (2006) The plant immune system. Nature 444(7117):323–329

Kandasamy K, Mohan SS, Raju R, Keerthikumar S, Kumar G, Venugopal A, Telikicherla D, Navarro JD, Mathivanan S, Pecquet C et al (2010) NetPath: a public resource of curated signal transduction pathways. Genome Biol 11(1):R3

Köhler S et al (2008) Walking the interactome for prioritization of candidate disease genes. Am J Hum Genet 82(4):949–958

Krallinger M et al (2009) PLAN2L: a web tool for integrated text mining and literature-based bioentity relation extraction. Nucleic Acids Res 37:W160–W165

Li P, Quan X, Jia G, Xiao J, Cloutier S, You FM (2016) RGAugury: a pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. BMC Genomics 17:852. https://doi.org/10.1186/s12864-016-3197-x

Liekens AM, De Knijf J, Daelemans W, Goethals B, De Rijk P, Del-Favero J (2011) BioGraph: unsupervised biomedical knowledge discovery via automated hypothesis generation. Genome Biol 12:R57

Makita Y, Kobayashi N, Mochizuki Y, Yoshida Y, Asano S, Heida N, Deshpande M, Bhatia R, Matsushima A, Ishii M, Kawaguchi S, Iida K, Hanada K, Kuromori T, Seki M, Shinozaki K, Toyoda T (2009) PosMed-plus: an intelligent search engine that inferentially integrates cross-species information resources for molecular breeding of plants. Plant Cell Physiol 50 (7):1249–1259

Martin GB (1999) Functional analysis of plant disease resistance genes and their downstream effectors. Curr Opin Plant Biol 2:273–279

Martin GB, Frary A, Wu TY, Brommon-Schenkel S, Chun-Wongse T, Earle ED, Tanksley SD (1993) Map based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262:1432–1436

Meyers BC, Kaushik S, Nandety RS (2005) Evolving disease resistance genes. Curr Opin Plant Biol 8:129–134

Michelmore R (2000) Genomic approaches to plant disease resistance. Curr Opin Plant Biol 3:125–131

Moreau Y, Tranchevent LC (2012) Computational tools for prioritizing candidate genes: boosting disease gene discovery. Nat Rev Genet 13:523–536

Morrison JL, Breitling R, Higham DJ, Gilbert DR (2005) GeneRank: using search engine technology for the analysis of microarray experiments. BMC Bioinformatics 6(1):233. https://doi.org/10.1186/1471-2105-6-233

Nitsch D, Tranchevent LC, Gonçalves JP, Vogt JK, Madeira SC, Moreau Y (2011) PINTA: a web server for network-based gene prioritization from expression data. Nucleic Acids Res 39:W334–W338. https://doi.org/10.1093/nar/gkr289

Parkinson H et al (2007) ArrayExpress—a public database of microarray experiments and gene expression profiles. Nucleic Acids Res 35:D747–D750

Parra G, Blanco E, Guigó R (2000) GeneID in drosophila. Genome Res 10(4):511–515

Patnala, Clements J, Batra J (2013) Candidate gene association studies: a comprehensive guide to useful in silico tools. BMC Genet 14:39

Rawat N (2016) Approaches for disease resistant candidate genes identification in plants: recent techniques and trends. Austin Food Sci 1(2):1010

Salamov AA, Solovyev VV (2000) Ab initio gene finding in Drosophila genomic DNA. Genome Res 10:516–522

Sanseverino W, Roma G, De Simone M, Faino L, Melito S, Stupka E, Frusciante L, Ercolano MR (2010a) PRGdb: a bioinformatics platform for plant resistance gene analysis. Nucleic Acids Res 38:D814–D821

Sanseverino W, Roma SGD, Faino M, Melito L, Stupka S, Ercolano MR (2010b) PRGdb: a bioinformatics platform for plant resistance gene analysis. Nucleic Acids Res 38(Database issue):D814–D821. https://doi.org/10.1093/nar/gkp978

Sarah ES, Le'on-Charles T, Sifrim A, Ardeshirdavani A, Davis J, Moreau Y (2015) Beegle: from literature mining to disease-gene discovery. Nucleic Acids Res. https://doi.org/10.1093/nar/gkv905

Schiex T, Moisan A, Rouze PE (2001) An eukaryotic gene finder that combines several sources of evidence. In: Sagot M (ed) Selected papers from JOBIM'2000, vol 2066. Springer, Verlag, pp 118–133

Seelow D, Schwarz JM, Schuelke M (2008) GeneDistiller—distilling candidate genes from linkage intervals. Ed Philip Awadalla. PLoS ONE 3(12):e3874

Sharma TR, Das A, Thakur S, Jalali BL (2014) Recent understanding on structure, function and evolution of plant disease resistance genes. Proc Indian Natn Sci Acad 80(1):83–93

Slota M, Maluszynski M, Szarejko I (2017) Bioinformatics-based assessment of the relevance of candidate genes for mutation discovery. In: Jankowicz-Cieslak J, Tai T, Kumlehn J, Till B (eds) Biotechnologies for plant mutation breeding. Springer, Cham, pp 263–280

Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res 33:W465–W467. http://bioinf.uni-greifswald.de/webaugustus/about.gsp

Stanke M, Keller O, Gunduz HA, Waack S, Morgenstern B (2006) AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res 34:W435–W439. https://doi.org/10.1093/nar/gkl200

Tamura K et al (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28 (10):2731–2739

Tamura K et al (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30(12):2725–2729

The Tomato Genome Consortium (TGC) (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485(7400):635–641

Tranchevent L-C, Ardeshirdavani A, ElShal S, Alcaide D, Aerts J, Auboeuf D et al (2016) Candidate gene prioritization with endeavour. Nucleic Acids Res 44:W117–W121

Van Der Biezen EA, Jones JDG (1998) Plant disease resistance proteins and gene-for-gene concept. Trends Biochem Sci 23(12):454–456

Waese J, Fan J, Pasha A, Yu H, Fucile G, Shi R, Cumming M, Kelley LA, Sternberg MJ, Krishnakumar V, Ferlanti E, Miller J, Town C, Stuerzlinger W, Provart NJ (2017) ePlant: visualizing and exploring multiple levels of data for hypothesis generation in plant biology. Plant Cell 29:1806–1821. https://doi.org/10.1105/tpc.17.00073

Wall L (1988) v13i001: Perl, a “replacement” for awk and sed, Part01/10. Newsgroup: comp. sources.unix

Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18(5):691–699. https://doi.org/10.1093/oxfordjournals.molbev.a003851

Yue P, Melamud E, Moult J (2006) SNPs3D: candidate gene and SNP selection for association studies. BMC Bioinforma 7:166

Zhang SL, LI D-f, Zhang G-s, Wang J-w, Niu N (2008) Prediction of Rice gene by Fgenesh. Sci Agric Sin 41(6):1567–1574. https://doi.org/10.3864/j.issn.0578-1752.2008.06.001

Zhang L, Li X, Tai J, Li W, Chen L (2012) Predicting candidate genes based on combined network topological features: a case study in coronary artery disease. PLoS One. https://doi.org/10.1371/journal.pone.0039542

Zimmermann P et al (2004) GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. Plant Physiol 136(1):2621–2632

Chapter 6
In Silico Approach in Tracing Persistence and Survival of *Rhizobium* in Soil

Konineeka Sen and Chandan Sengupta

Abstract Efforts to unravel the survival mechanism of *Rhizobium* in soil by successful nodulation of legume roots and corresponding enhancement of their N_2 fixing ability in soil have been initiated as early as 1950s. Although such studies were at its infancy for over three decades, most of the works were confined to in vitro studies carried out in simulations mimicking field conditions. Ongoing researches since the 1980s have been concentrating on an array of minute and intense parameters like soil texture, moisture, temperature antibiotic sensitivity, particulate size of soil aggregates and adverse conditions of flooding and drought, respectively, exploring the prospects of soil amendments with inert fillers. Nowadays with the ever-increasing demand for food production and ensuring food security to the exploding global population, rhizobia are being introduced more in soil milieu in association with arbuscular mycorrhizal fungi, plant growth-promoting *Rhizobacteria* to improve crop growth and yield. To counteract inhibitory effects of antagonistic microbes in soil, plant roots are being infected with rhizobial inoculants in pyrolyzed biomass carriers. This chapter focuses on molecular tools to gain insight into the mechanism of overcoming heat stress and desiccation by expression of heat shock proteins and blue light sensitivity on nodulation that has added new dimensions to this area of research.

6.1 Introduction

The survival of *Bradyrhizobium japonicum* in soil nodulating soybean 20 years after inoculation even though soybean was not grown in the soil 17 years after introduction of the inoculants was reported in the polish soil a couple of years back (Narozna et al. 2015). *Rhizobium* has been found to survive 10–125 years in soil after

K. Sen (✉)
Department of Botany, Pingla Thana Mahavidyalaya, Maligram, India

C. Sengupta
Microbiology Division, Department of Botany, University of Kalyani, Kalyani, India

D. K. Choudhary et al. (eds.), *In Silico Approach for Sustainable Agriculture*,
https://doi.org/10.1007/978-981-13-0347-0_6

cultivation of their respective legume host. The symbiotic relationship of legumes with the root-nodule bacteria, *Rhizobium*, that fix atmospheric N_2 for such plants came to light only in the early twentieth century (Fred et al. 1932) although legumes have long been recognised as crops capable of improving soil fertility as early as in the late nineteenth century. Legumes are best suited to the tropical agriculture, but their failure in fixing atmospheric N_2 is largely due to the factors that there are few or no rhizobia in the soil, *Rhizobium* living in soil are incapable of nodulating such legumes or the indigenous bacterial strains are not effective N_2 fixers with the legume under cultivation. The commercially introduced rhizobial inocula do not contain the appropriate symbiont; they do not survive for long to infect the first crop or the subsequent crops.

Nutman observed that in soil where wheat had been grown for almost half a century, since 1863, serial dilution of such soil yielded most abundantly *Rhizobium leguminosarum* about 28,000/g of dry soil with *R. trifolii* next nearing it about 310/g of soil that could effectively nodulate and fix N_2 as well. *R. lupini and R. meliloti* were found relatively sparsely and sporadically in some fields, averaging 310/g and 10/g, respectively. Nutman and Ross (1970) observed fields continuously cultivated from 1863 but left fallow for 8 years or so, the number of *Rhizobium trifolii* fluctuated between 1000 and 100,000/g and *R. leguminosarum* varied between 10 and 1000/g of soil. Earlier in another study, Norman (1942) observed that the soybean yield or nodulation in less fertile sandy loam as well as more fertile loam did not decline significantly even after 24 years after soybean was last planted to such soil. Weaver and Frederick concluded that there is no correlation between the number of *Rhizobium* present in soil and the number of years since the legumes were last cultivated in such soils. In contrast, Elkms et al. opined that the number of *R. japonicum* decreases in absence of compatible hosts. This chapter will focus on the correlation between population fluctuation of *Rhizobium* spp. and the various soil parameters that affect growth and persistence of rhizobia in soil, to get an insight into the mechanism of survival of rhizobia in soil. However, since such studies have initiated as early as mid of twentieth century, recent developments pertaining to molecular approach in tracking such mechanism have been highlighted.

6.2 Temperature Tolerance of *Rhizobium* in Soil

Despite of an early report of *Rhizobium* to survive over a wide range of temperature 0–50 °C (Fred et al. 1932), now it has been well established that *Rhizobium* strains are susceptible to elevated temperatures beyond 60 °C (Vincent 1977) but survive better at low temperatures. Although Graham and Parker (1964) reported 22 isolates out of 24 could survive a temperature of 50 °C, this was contradicted by Sengupta (1988) who observed that two strains of rice bean *Rhizobium* and *R. leguminosarum* were killed by heating at 50 °C for 20 min. In contrast, the same study revealed survival of two marker strains, viz., *R. leguminosarum* (fast grower) and cowpea *Rhizobium* (slow-grower) in soil, the population increased exponentially during first

week of incubation irrespective of temperature conditions, and this growth continued for next 5 weeks. In a survey of temperature tolerance for growth of temperate and tropical strains of *Rhizobium*, Bowen and Kennedy (1959) opined that temperate strains were more tolerant to a wide range of temperature varying between 36.5 and 42.5 °C; on the average *R. meliloti* were more tolerant than *R. leguminosarum* and *R. trifoli.* On the other hand, tropical strains could withstand a temperature of 32–40 °C. Mahler and Wollum (1981) based on the growth response of 35 strains *R. japonicum* in liquid medium found that tolerance to high temperature was not related to their geographical origin, whereas Wilkins observed that rhizobia strains associated with *Acacia*, *Lotus* and *Psoralea* from warmer soils of New South Wales survived higher temperatures than strains obtained from cooler regions of New England Tableland, thus contradicting the ecological adaptation of such strains. She could not, however, generalise the same for strains from *Medicago* sp. growing in warmer and cooler soils, respectively. Cowpea rhizobia isolated from hot dry environment of Savannah, West Africa, showed similar results (Eaglesham et al. 1981). All isolates were tolerant to high temperature and showed remarkable growth in yeast extract mannitol at 37 °C, but isolates growing in more humid areas of West Africa could not resist such high temperature (Eaglersham and Ayabana 1984).

Several reports of better survival of rhizobia at temperatures above 40 °C in dry soil than in moist soils have been obtained (Bowen and Kennedy 1959; Foulds 1971). Strains of *R. meliloti*, *R. japonicum* and *R. lupini survived for* 5–6 h at even 70 °C in dry sandy soils. Large number of strains of *Rhizobium* from pea, clover, lucerne and tropical legumes, however, declined rapidly in moist soils only at 40 °C (Bowen and Kennedy 1959). Willatt and Tighe observed that in Rhodesia, soil temperature under soybeans in unshaded condition during February–March was high enough to inhibit rhizobia growth, although nodulation of legumes occurred.

Tolerance to cold, heat and heat shock of several species of *Mesorhizobia* was evaluated in different province of Portugal (Alexandre and Oliviera 2010). A close association could be drawn between isolate phenotype and the province of origin. In an analysis of *dnaK* and *groESL* expression by northern hybridisation, using isolates of three species, an increase in transcript levels with heat but not with cold stress was reported. Thus it was inferred that higher induction of chaperone genes enabled such strains to better adapt to heat tolerance.

6.3 Survival Sensitivity of Rhizobia to Soil Drying

Several authors have reported that rhizobia are markedly sensitive to soil drying. Vincent et al. (1962) observed that the population of *R. trifolii* declined significantly during drying of seeds. In a similar study, Marshall (1964) noted the survival of root-nodule bacteria in dry sandy soil exposed to high temperature. While the fast-growing strains, viz. *R. trifolii* and *R meliloti*, failed to survive, the slow-growing strains, *R. japonicum* and *R. lupini*, could resist the effects of high temperature and desiccation in sandy soil. Such observations were reinforced by a later study by

Bushby and Marshall (1977) where nine slow-growing rhizobia exhibited 0.6–7% survival after overnight drying, whereas such nine fast growers showed only 0.02–2.5% survival. They suggested that the decline in rhizobial population during desiccation was affected by soil type and clay content.

High soil temperature limits nodule formation, and dinitrogen fixation was studied in bean-nodulating *Rhizobium* strains. Acetylene reduction activity of nodulated bean roots was strongly reduced at 35–40 °C in plants nodulated by strains CIAT899 and CNPAF512. Such strains tested under free-living condition, survival at 40 °C, kinetics of methionine uptake and protein synthesis at the temperature range of 35–40 °C were optimum. Synthesis of heat shock proteins, detected in both strains, 14 heat shock proteins of CNPAF512 and 6 of CIAT899 accounts for the thermotolerance. Strain CIAT899 could even resist a temperature of 45–50 °C which is also in agreement with the data obtained by Wang and Martinez Romero (2000) who witnessed that *R. tropici* is more thermotolerant than *R. leguminosarum* bv. phaseoli. Protein synthesis with the increase in temperature from 40 to 45 °C was hampered in CNPAF512 after 60 min as indicated by accumulation of S^{35} methionine in contrast to CIAT899 where cells accumulated the amino acid at constant rate. Upon synthesis of specific proteins investigated by pulse labelling with L [^{35} S] methionine and analysing by 1D and 2D gel electrophoresis, a couple of isoforms of 7 proteins varying in molecular mass between 19 and 97 kDa were obtained with increase in temperature from 40 to 45 °C.

The failure of nodulation under elevated temperature is possibly due to the effects of temperature on the root of the host plant. Lee and Dobereiner (1982) pointed out strains showing high-temperature tolerance in pure culture might not show similar performance in soil and vice versa. Temperature tolerance should be correlated with nitrogenase activity before selecting a temperature-resistant strain under field condition, although strains show heat tolerance but are mostly sensitive to moist heat.

6.4 Effect of Desiccation on Survival

Rhizobia tend to die rapidly under drying conditions. Vincent (1977) reported that rhizobial cells were killed rapidly when dried at normal atmospheric pressure, death being most rapid during the first 24–27 h. However, some sugars like maltose and sucrose when present in the medium could prevent death of cells due to desiccation. The length and severity of dry periods and temperature variations are more in the tropics. The survival of rhizobia in the tropical and temperate soils has been extensively studied by Lowendorf (1980) and Eaglesham and Ayanaba (1984). The ability of rhizobia to survive desiccation depends on their ability to cope with radiation stress, reactive oxygen species, certain salts and solutes and temperature extremes (Potts 1994; Ramos et al. 2001). Desiccation stress can be differentiated into three phases: drying (stage I), storage (stage II) and rewetting (stage III). The severity and rate of drying as well as rewetting and the storage period profoundly influence desiccation. The effects of desiccation are fourfold: (i) the accumulation of

salts and solutes, (ii) hyperosmotic stress, (iii) impairment of metabolism when critical water level has been reached and (iv) accumulation of damage when aqueous monolayer has been removed from macromolecules. The damage during storage is comparable to damage due to ionising radiation and UV radiation and damage due to ROS.

6.4.1 *Influence of Soil Types on Survival of Temperature Extremes and Desiccation by Rhizobia*

Marshall and Roberts (1963) first highlighted that the performance of fast-growing strains of rhizobia under high temperatures could be improved by amending the soils with materials having fine particle sizes such as fly ash and clay. Marshall (1964) provided further evidence that soil particles protect rhizobia from environmental stress of high temperature and desiccation. Bushby and Marshall demonstrated that amendment of sandy soil with montmorillonite could protect the fast-growing rhizobia from desiccation but not the slow-growing ones, whereas kaolinite gave little or no protection to either group. Bushby and Marshall (1977) opined that montmorillonite protects rhizobia by reducing water content of cells. Jansen Van Rensburg and Strijdom (1980) confirmed the observations of the former in a separate study of the survival of fast- and slow-growing strains of rhizobia under relatively mild desiccation. Slow-growing rhizobia were more resistant to severe desiccation than the fast-growing ones. Further, they supported the mechanism of survival of slow-growing strains to desiccation as proposed by Bushby and Marshall (1977). However, Pena-Cabriales and Alexander (1979) using four fast-growing and three slow-growing strains of rhizobia contradicted the findings of Marshall and his associates. In a silt loam soil undergoing desiccation for 10 days, the survival rates were 0.39–0.92% and 0.39–0.47% for fast and slow growers, respectively. This discrepancy between the two results might be due to soil type used by these authors. While former workers used sandy soil, the latter group used silt loam.

Osa-Afiana and Alexander (1982) further reported that the population of two strains of *R. japonicum* and two strains of cowpea *Rhizobium* both of slow-growing type declined drastically in soil undergoing drying: the extent of drying however depended on strain and soil type used. Decline of most of these strains was more when they were grown on glass beads than on fine sand. Several soil clay minerals were tested for their ability to protect against desiccation. Illite and kaolinite had little or no protective effects. Montmorillonite, however, offered some protection to the two strains but not to the other two more desiccation-sensitive strains on fine sand. Further studies have shown that the survival of three strains of *R. japonicum* during drying in fine sand depended on the relative humidity of the ambient atmosphere and survival increased with the decrease in relative humidity. These studies were in confirmation with the work of Bushby and Marshall (1977).

6.5 Effect of Salt and Osmotic Stress on the Survival Response of Rhizobia

Osmotic stress and desiccation stress are the two complementary phases where osmotic phase refers to the abundance of solutes, while desiccation stress results from lack of water. Chen and Alexander established a correlation between the two stresses as he observed that the isolates that have low water activity can better survive desiccation. The presence of NaCl decreases chances of survival of rhizobia in soil, seed inocula and in vitro when cells are resuspended in water containing 400 mM NaCl prior to drying (Kosanke et al. 1999). Contrastingly *Sinorhizobium meliloti* strain RCR2011 showed an increase in survival during desiccation, while other strains of *S. meliloti* did not show such similar results (Vriezen et al. 2006). This could be accounted to the addition of sulphate salts prior to drying that leads to an elevated survival rather than after addition of chlorides, despite a decrease in osmotic tension of the medium. Hence it implies that ionic strength and toxicity induce differential physiological responses to the several salts and are more important in survival than osmotic pressure or water activity (Vriezen et al. 2007). It may be concluded that an overlap in response to osmotic and salt stress and ability to survive desiccation exists as depicted in Fig. 6.1, but it is quite unlikely to be limited to the accumulation of desiccation protectant.

6.5.1 Response of Rhizobia on Treatment with NaCl

Strains of rhizobia differ in their response to increase in salt stress along with osmotic pressure, and they resemble the enteric bacteria in such responses.

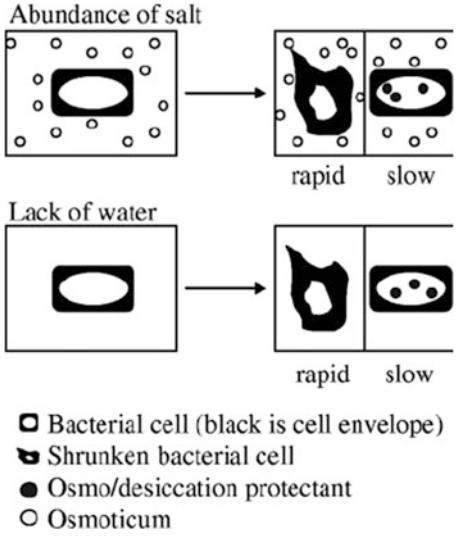


Fig. 6.1 Physiological response of *Sinorhizobium* to desiccation and osmotic stress showing an overlap in response (Reproduced from Vriezen et al. 2007)

Generally, after an osmotic upshift, metabolism slows down. This can be supported by the findings that genes involved in the tricarboxylic acid cycle, in the uptake of carbon source, in respiratory chains are repressed. Astonishingly most of the genes that are downregulated by NaCl encode ribosomal proteins.

The genetic mechanism underlying survival of *Sinorhizobium* after desiccation in presence of NaCl has been unravelled by Vriezen et al. Since salt stress is accompanied by loss of water activity, identification of NaCl-induced loci has enabled to correlate the survival mechanism involved in desiccation. This approach enabled identification of loci asn O and ngg that by their reduced ability to grow in higher sodium chloride due to their inability to produce the osmoprotectant, N-acetylglutaminylglutamine (NAGGN). The amino acid sequence similarity with enzyme producing β lactam inhibitor as well as increased resistance to β lactam antibiotics by asn O indicates that this locus is involved in production of such inhibitors.

Rhizobia are capable of accumulating K^+ ions that do not require any new protein synthesis suggesting that the K^+ ion uptake is regulated biochemically and is used as a secondary messenger. Nogales et al. had reported such a high-affinity K^+ uptake (kup) system in *Rhizobium tropici*, and a corresponding homolog was found in *S. meliloti* 1021.

6.6 Soil Moisture Content on Survival

Hedlin and Newton first studied the effect of soil moisture content on survival of *R. leguminosarum* in Canadian soil. The population increased about ten times within 38 days, but gradually as the water content decreased, the population declined and became nil by about 148 days. Prior to this, earlier studies failed to establish a correlation between soil moisture content and the population fluctuations of species of *Rhizobium*, viz. *R. meliloti*, *R. leguminosarum* and *R. trifolii* in silt loam of Washington (Vandecaveye 1927). Osa-Afiana and Alexander (1979) discovered that a critical water level of 10% was crucial for population increase, but further increase in moisture content to 22%, 35% and 45% proved to be detrimental to their growth.

6.7 Effect of Flooding on Survival

The effect of flooding of soil on the population dynamics of *R. leguminosarum* cells was studied by Boonkird and Vadeesirisal. Soil samples were inoculated with marker cells, and sterilised water was added so that the water level above the soil surface was 2, 4 and 6 cm in separate bottles. The number of surviving cells at specific intervals was recorded. Control soil was adjusted to 60% water holding capacity (WHC). Flooding of soil had a significant effect on survival of cells. Despite inundation of soil under different depth of water levels, the population of *R. leguminosarum* kept increasing during the first 2 weeks of incubation. Amazingly the increase in sterilised soil under flooding condition was however significantly less

Table 6.1 Effect of NaCl on the mean doubling time (MDT) of 11 strains of *Rhizobium*

	Mean doubling time			
	0 mM NaCl	50 mM NaCl	100 mM NaCl	
Strain	(1.2 mmhos)	(6.7 mmhos)	(13.1 mmhos)	Relative tolerance[c]
17E[a]	2.7	5.3	8.4	3.1
Hawaii 5-0	3.9	11.8	N.G.[b]	–
7B[a]	6.0	7.7	69.7	11.6
Web 48	8.7	11.9	N.G.	–
TAL 425	7.4	7.5	17.4	2.4
23B[a]	6.2	7.9	10.1	1.6
8a	5.9	7.1	N.G.	–
14E[a]	7.4	7.8	69.0	9.3
TAL 426	8.4	7.5	26.4	3.1
21A[b]	10.3	12.6	17.6	1.7
USDA 110	5.5	6.3	27.2	5.0

Table reproduced from thesis of Singleton (1982)
[a]Isolated from salt-affected soil
[b]No growth
[c]The ratio of salt-affected soils. MDT at 13.1 mmhos cm^{-1}: MDT at 1.2 mmhos cm^{-1}

than the non-sterile soil. Apparently the flooding of the soil retarded growth and survival of *R. leguminosarum* in sterile as well as non-sterile soils. The decrease in population in non-sterile soil was significant relative to control (Table 6.1). Simulations carried out under field conditions showed that after 3 weeks of flooding, the rhizobial population decreased to half approximately, but a population density of 1000cells/g of soil could be maintained throughout the cultivation season of paddy for almost all the study areas. This number was estimated to be adequate for nodulation of peanut.

Since previous experiments revealed the inhibitory effects of flooding on population dynamics of rhizobia, efforts were taken to overcome such effects by soil amendment with organic matter. The distribution pattern of surviving cells in soil and water phase was investigated (Sengupta 1988). From the results it was evident that the population of the two rhizobial species tested decreased over time under flooding condition even when the soil was amended with organic matter. For soil amended with sucrose, the decrease in population for both the species was more rapid than non-amended soil. In contrast, amendment of soil with mannitol, however, stimulated the growth of both the *Rhizobium* species during the first 2 weeks but declined thereafter.

A small fraction of rhizobial cells was noted to thrive in the water-soil phase of the control. Further the distribution pattern in the soil and water phase was not significantly altered by amendment of soil. In soils amended with sucrose, the decline in population was significant both in soil and water phase. In comparison to amendment with mannitol, the population in both soil and water phase was higher than the control soil during the first 2 weeks.

Interestingly, it was noted that the population of soil bacteria and fungi in soil and water phase of control soil was estimated to exist in the proportion of 2:1. In case of

soil inoculation with *R. leguminosarum*, only a small number of soil bacteria and fungi were distributed in water phase, and subsequent amendment of soil with sucrose and mannitol increased the number in soil phase.

6.8 Effect of Light on Population Dynamics of Rhizobia

The effect of light on survival of rhizobial cells was investigated for non-sterile and sterile soils, incubating under different light conditions such as continuous darkness, continuous fluorescent light, diffused and sunlight (Ballhorn et al. 2016). In sterilised soils, continuous darkness allowed best survival. Subsequently, none of the other light could remarkably decrease proliferation of rhizobial cells, except direct sunlight. However, exposure to direct sunlight caused significant decrease of *R. leguminosarum* population approximately after 21 days of incubation. After 56 days, the population growth in sterilised soil was stalled to an extent comparable to non-sterilised soil. Unlike others, rice bean *Rhizobium* exhibited relatively lower decrease in population on exposure to direct sunlight.

When incubated to non-sterilised soils, light conditions could hardly influence survival of test isolates. Likewise direct sunlight had profound inhibitory effects on the population dynamics of test isolates.

In a recent study, it was found that blue light perception by legume roots inoculated with *Mesorhizobium loti* was detected to inhibit root nodulation in *Lotus meliloti* (Shimomura et al. 2016). In this experiment, using RNA interference, the expression of phototropin and cryptochrome genes was suppressed in *L. japonicus* hairy roots. Under the influence of blue light, plants transformed with empty vector did not develop nodules, whereas plants expressing suppression of cry1 and cry2 genes were capable of forming nodules. Subsequently the effect of blue light on rhizobial growth was also measured to ascertain the role of blue light on reduction of population. Consequently it was revealed that blue light had remarkable inhibitory influence, although red light proved to be ineffective in reducing such population. Growth under blue light could be partially revived in mutants in which LOV-HK/PAS- and photolyase-related genes were suppressed. However, when Ljcry1A- and Ljcry2B- silenced plants were inoculated with mutant strains, nodulation increased cumulatively.

6.9 Effect of Soil pH on Survival

Influence of soil pH on growth and survival of rhizobia were tested on loam and sandy loam soil samples varying in different pH values (Niste et al. 2013). Simulation of field conditions was replicated by incubating soil samples inoculated with mutant strains of *Rhizobium* in glass jars for 2 months. The population of surviving cells were determined by plate count method. From the results, it was apparent that the population of both rice bean *Rhizobium* and *R. leguminosarum* rapidly

deteriorated in non-sterile soil irrespective of its pH being acidic, neutral or alkaline in nature. Considering the exponential growth of such population and the number of surviving cells in soil varying in pH, it could be inferred that the rate of decline in acidic soil over time was even more rapid compared to alkaline or neutral soils in case of *R. leguminosarum*. Contrary to this finding, alkaline and neutral soils were not found to be inhibitory to the growth of such bacteria.

In non-sterile soils having different pH levels, the decline in population could be accounted to either pH effect, microbial interaction or under dual influence. In contrary to non-sterile soils, having pH value of 6.9, both the test rhizobial species increased in population and such a high density persisted throughout the incubation period, whereas at a pH value of 4.5, the population of both the *Rhizobium* species increased slightly during the first 2 weeks and then declined slowly to attain the native population count. In case of rice bean *Rhizobium*, the decline occurred after 8 weeks of incubation. Contrastingly for *R. leguminosarum*, the population declined to 50% level after 8 weeks.

6.10 Salt Tolerance of Rhizobia

The salt tolerance of different symbionts in the different *Rhizobium*-legume symbiosis differs from one another. Though there exists some variation in salt tolerance of legumes, none of the agriculture legumes have been reported to be significantly salt tolerant. The growth of all rhizobia decreases with the slightest increase in salt concentration and was demonstrated by Singleton (1982), but to the contrary Pillai and Sen (1973) showed that the growth rate of *Rhizobium* increases relatively with addition of 1% NaCl in growth media (EC = 18 mmhos cm^{-1}). A couple of years later, Steinhorne and Roughley observed that the growth rate of *R. trifolii* and *R. meliloti* was retarded with the addition of salt to growth media, thus supporting previous reports. The relative tolerance of different strains to different salt concentrations of *Rhizobium* varies which could be measured by calculating mean doubling time (MDT), tabulated by Singleton 1982 (Table 6.1). Not much increase in mean doubling time was observed until the electrical conductivity was raised to 13.1mmhos which is equivalent to 28% seawater.

There is an inverse relationship between soil moisture tension and salinity in microenvironments. With the reduction in soil moisture content, the salt concentration increases. To test the effect of soil moisture tension and salinity on the survival of *Rhizobium*, treatments involving conditions optimum for plant growth (EC = 0.2 mmhos cm^{-1}) as well as extremes that are not favourable or inadequate for plant growth were done. None of the four strains of *Rhizobium* tested lost viability even under extreme conditions. *R. leguminosarum* strain Hawaii 5-0, USDA 110 and 21A that showed to be sensitive to salt in both growth rate study and growth in salt solution showed very slow growth but survived best in clay soil. 21A showed good tolerance to soil in solution with EC of 43.0 mmhos cm^{-1}. Gradually with time the strain showed loss in viability under the combined influence of high salinity and desiccation. The decline was not absolutely due to low moisture

content as 21A could thrive at −15 bars and 5.0 mmhos cms^{-1}. *R. japonicum* strain USDA 110 probably lost viability as a function of increasing stress from osmotic and matric component of soil water potential. The fast-growing isolate 17 E was resistant to all the stress components.

In separate studies, Marshall (1964) and Bushby and Marshall (1977) observed that slow-growing rhizobia could survive best in desiccated sandy soil than the fast-growing strains. Mahler and Wollum (1981) showed that even clay soil at −15 bars, moisture tension retarded growth of many strains of *R. japonicum* and *R. leguminosarum*.

6.11 Influence of Particle Size of Soil Aggregates on Distribution of *Rhizobium*

The distribution of *R. leguminosarum* population across different size classes of soil aggregates, with particle sizes ranging from 0.25 to 5.0 mm under the effect of winter cover crop, was studied using a combination of plant infection-soil dilution technique, most probable number (MPN) and immunofluorescence direct count (IFDC) microscopy (Mendes and Bottomley 1998). For the summer crop treatment, either fallow land or plots covered with red clover crop and triticale were surveyed from April to September. The rhizobia population were heterogeneously distributed over the different size classes of soil aggregates, and the distribution was influenced by cover crop treatment and sampling time, respectively. For the September sampling, the smallest size class aggregate (0.25 mm) recovered from red clover plots, carried between 30% and 70% of the nodulating *R. leguminosarum* population, while for the June sampling, only 6% of the population could be recovered. IFDC microscopy for the June sampling revealed that 1.0–2.0 mm size class of aggregates from the red clover cover treatment carried marked increase in population density than the aggregate size class of 0.5, 2.0–5.0 mm), whereas in the September sampling, significant population density was noted for size classes of soil aggregates between 0.25 and 0.5 mm, rather than for aggregates ranging between >1.0 and <0.25 mm. Similar trends of distribution were noted for two other serotypes of *Rhizobium*, viz. (AR6 and AS36) for the June and September sampling, respectively. From this study it was evident that the existence of structural microsites determines the suitability of the soil to support growth and proliferation of bacteria, and such effects are augmented by the presence and type of plant cover in the soil.

6.12 Conclusion

In a nutshell the salient findings of this study may be summarised as follows:

The study on survival rate of the test strains of rhizobia in sterilised soil showed that the multiplication was constant during the tenure of the experimental period

without showing least decline, although in non-sterilised soil the population decline is drastic and within 8 weeks becomes 1/10 or 1/100 of the initial population. This clearly indicates the influence of soil microflora on the population density and persistence of introduced rhizobial strain.

Temperature sensitivity of the strains of *Rhizobium* was assessed to be above 50 °C. At this temperature 20 min heating could rapidly kill cells of *R. leguminosarum* and rice bean *Rhizobium*. Likewise, temperature tolerance has been rated to be quite high, and hence such strains are flexible to grow under varied climatic conditions and different ecosystems. Moreover, the fast- and the slow-growing strains are comparable with respect to their temperature tolerance. Temperature tolerance of chickpea rhizobia has been attributed to the expression of chaperone genes such as dnaKj and groESL (Alexandre and Oliviera 2010). Isolates exhibited better growth at higher temperature in dry soil (10% WHC) compared to moist soil (100% WHC) where the population decline was recorded to be 10- to 100-fold more in comparison to dry soil, although the population decline in non-sterile soil was relatively higher but is not completely lost at even 45 °C.

Review on the effect of desiccation on survival of *Rhizobium* has revealed two conflicting views. Several reports support the view that population decline is witnessed when soils are dried. In contrast, evidences suggesting rhizobia can withstand desiccation for long periods in air dried soil have been opined by others. Desiccation and salinity have been reported to limit N_2 fixation, thus justifying the former school of opinion.

Soil moisture content conducive for supporting growth and persistence of rhizobia in soil varied between 40% and 60% WHC, but no significant difference in survival pattern was found for the slow- and fast-growing species. Survival was significantly better for sterile soil compared to non-sterile soil.

Flooding of soil adversely affected population of *R. leguminosarum* in non-sterile soil as well as sterile soil, though such effects were more pronounced for non-sterile soil. Population decrease in flooded soil can be well correlated with the depth of free water above the soil level. Subsequently, water level of 2–4 cm above the soil was conducive for the persistence of bacteria, and the decline was moderate (Verma and Rawat 1991). Saturation of 60% WHC was found to be adequate to sustain optimum *Rhizobium* density in soil. Supplementing soil with 0.5% farm yard manure could support ample proliferation of the bacteria in soil. Soil moisture level and organic C interacted leading to the depreciation of C in simulated sandy biomes.

Rhizobium spp. growing in various habitats and temperature zones can withstand acidic pH range of 4.5–6.9. Although alkaline and neutral pH are preferred by rhizobia, the rate of decline of such populations is higher in acidic non-sterile soil than sterile soil. Salt tolerance of rice bean *Rhizobium* and *R. leguminosarum* was estimated to range between EC equivalent of 13 mmhos cm^{-1} and 43 mmhos cm^{-1}.

The potential contribution of *Rhizobium* inoculants in remediation of fallow, low fertility and waste lands should not be underestimated. From the present study, it may be recommended that temperature range of 36–42 °C and a critical temperature of 50 °C, critical water level of 10%, flooding water levels of 2–4 cm above the soil (WHC = 60%) and particulate size of soil aggregates varying between 1 and 2 mm

preferably sandy soil can easily withstand by strains of *Rhizobium* and hence optimum for exploiting such inoculants under varied environmental conditions for improving growth and yield of crops. Slow-growing strains like *R. japonicum* and *R. lupini* can resist high temperatures and are congenial for desert and extreme habitats.

Acknowledgment We are deeply indebted to Prof. Kanak Ranjan Samaddar, Retired faculty of Plant Pathology Division, Department of Botany, University of Kalyani and a renowned researcher in the field of Soil Ecology for pioneering such work in India and contemplating such manuscript as presented here.

References

Alexandre A, Oliviera S (2010) Most heat tolerant rhizobia shows high induction of major chaperone genes upon stress upon stress. FEMS Microbiol Ecol 75:28–36

Ballhorn DJ, Schadler M, Elias JD, Millar JA, Kautz S (2016) Friend or foe- light availability determines the relationship between mycorrhizal fungi, rhizobia and lima bean (*Phaseolus lunatus* L.). PLoS One 11(5):e0154116. https://doi.org/10.1371/journalpone

Bowen GD, Kennedy MM (1959) Effect of high soil temperatures of *Rhizobium* spp. Queensland J Agric Sci 16:177–197

Bushby HVA, Marshall KC (1977) Water status of rhizobia in relation to their susceptibility to desiccation and to their protection by montmorillonite. J Gen Microbiol 99:12–27

Eaglersham ARJ, Ayabana A (1984) Tropical stress ecology of rhizobia, root nodulation and legume fixation. In: Subba Rao NS (ed) Current development in biological nitrogen fixation. Oxford and IBH Publishing Co, New Delhi, pp 1–35

Eaglesham A, Seaman B, Ahmad H, Hassouna S, Ayabana A, Mulongoy K (1981) High temperature tolerant cow-pea rhizobia. In: Gibson AH, Newton WE (eds) Current perspectives in nitrogen fixation. Australian Academy of Sciences, Canberra, p 436

Foulds W (1971) Effects of drought on three species of rhizobium. Plant Soil 35:665–667

Fred EB, Baldwin IR, Mc Coy E (1932) Root nodule bacteria and leguminous plants. University of Wisconsin Studies in Science, Madison

Graham PH, Parker CA (1964) Diagnostic features in the characterization of the root nodule bacteria of legumes. Plant Soil 20:383–396

Jansen Van Rensburg H, Strijdom WB (1980) Survival of slow and fast-growing *Rhizobium* spp. under conditions of relatively mild desiccation. Soil Biol Biochem 12:353–356

Kosanke JW, Osburn RM, Smith RS, Lipha Tech Inc. (1999) Process for preparation of bacterial agricultural products. Canadian patent 2073507

Lee K-K, Dobereiner J (1982) Effects of excessive temperatures on rhizobia growth nodulation. Pasqui Agro Bras 17:181–184

Lowendorf HS (1980) Factors affecting survival of *Rhizobium* in soil. Adv Microb Ecol 4:87–124

Mahler, Wollum AG (1981) Growth of *Rhizobium japonicum* strains at temperatures above 27° C. Appl Environ Microbiol 42:272–276

Marshall KC (1964) Survival of root nodule bacteria in dry soils exposed to high temperatures. Aust J Agric Res 15:273–281

Marshall KC, Roberts FJ (1963) Influence of fine particle materials on survival of *Rhizobium trifolii* in sandy soil. Nature 198:410–411

Mendes IC, Bottomley PJ (1998) Distribution of population of *Rhizobium leguminosarum* bv trifolii among different size classes of soil aggregates. Appl Environ Microbiol 64(3):970–975

Narozna D, Pudelko K, Koliczak J, Golinska B, Sugawara M, Madrzak CJ, Sadowsky MJ (2015) Survival and competitiveness of Bradyrhizobium japonicum strains 20 years after introduction into field locations in Poland. Appl Environ Microbiol 81(16):5552–5559

Niste M, Vidican R, Rotar I, Pop R (2013) The effect of pH stress on the survival of *Rhizobium trifolii* and *Sinorhizobium meliloti* in vitro. Bull UASMV Serie Agric 70(2):449–450

Nutman PS, Ross GJS (1970) Rhizobium in the soils of Rothamsted and Woburn farms. Ann Rep Rothamsted Exp Stat for 1969, Part 2: 148–167

Osa-Afiana LO, Alexander M (1979) Effect of moisture on the survival of *Rhizobium* in soil. Soil Sci Soc Am J 43:925–930

Osa-Afiana LO, Alexander M (1982) Clay and the survival of *Rhizobium in soil during desiccation*. Soil Sci Soc Am J 46:285–288

Pena-Cabriales JJ, Alexander M (1979) Survival of *Rhizobium* in soils undergoing drying. Soil Sci Soc Am J 43:962–966

Pillai RN, Sen A (1973) Salt tolerance of *Rhizobium* from *Dolichos lablab*. Zbl Bakt Abt II 128:538–542

Potts M (1994) Desiccation tolerance of prokaryotes. Microbiol Rev 58:755–805

Ramos JL, Gallegos MT, Marques M, Ramos- Gonzalez MI, Espinosa- Urgel M, Segura A (2001) Responses of gram negative bacteria to certain environmental stressors. Curr Opin Microbiol 4:166–171

Sengupta C (1988) Ecology of survival of *Rhizobium* in soil. Ph D thesis, university of Kalyani, 270 pp

Shimomura A, Naka A, Miyazaki N, Moruichi S, Arima S, Sato S, Hirakawa H, Hayashi M, Mzaymon M, Hirsch AM, Suzuki A (2016) Blue light perception by both roots and rhizobia inhibits nodule formation in *Lotus japonicas*. Mol Plant-Microbe Interact 29(10):786–796

Singleton PW (1982) The effect of salinity on *Rhizobium* survival, nodule function and nodule formation in the soybean *Rhizobium japonicum* symbiosis. Ph. D thesis, University of Hawaii

Vandecaveye (1927) Effect of moisture, temperature and other climatic conditions on *R. leguminosarum* in the soil. Soil Sci 23:355–362

Verma LN, Rawat AK (1991) Survival of *Rhizobium japonicum* as affected by different levels of moisture and organic matter in sandy soils. Zentralblatt fur Microbiologie 146(7):545–547

Vincent JM (1977) Rhizobium, general microbiology. In: Hardy RWF, Silver WS (eds) A treatise on dinitrogen fixation. Section III: biology. Wiley, New York, pp 277–366

Vincent JM, Thompson JA, Donovan KO (1962) Death of root nodule bacteria on drying. Aust J Agric Res 13:258–270

Vriezen JAC, de Brujn FJ, Nuslein K (2006) Desiccation responses of survival of *Sinorhizobium meliloti* USDA 1021 in relation to growth phase, temperature, chloride and sulphate availability. Lett AppliedMicrobiol 42:172–178

Vriezen JAC, de Brujn FJ, Nuslein K (2007) Responses of rhizobia to desiccation in relation to osmotic stress, oxygen and temperature. Appl Environ Microbiol 73:3451–3459

Wang ET, Romero M (2000) Sesbanea herbacea- Rhizobium huautlense nodulation in flooded soils and comparative characterization of S. Herbacea nodulating rhizobia in different environments. Microb Ecol 40(1):25–32

Chapter 7
Bradyrhizobia-Mediated Drought Tolerance in Soybean and Mechanisms Involved

Abhishek Bharti, Richa Agnihotri, Hemant S. Maheshwari, Anil Prakash, and Mahaveer P. Sharma

Abstract Among various legume crops grown across the world, soybean (*Glycine max* L. Merrill) is emerging as one of the fastest-growing oilseed crops in the world containing about 40% protein and 20% oil and is also being used as potential feed for animals. Drought stress being a major abiotic stress factor affects the soybean productivity adversely. Thus, to enhance the productivity of soybean, besides managing the nutrients, stress management is of utmost importance. There is enormous potential for opportunity of application of microbes especially bradyrhizobia either applied alone or co-inoculated with plant growth-promoting rhizobacteria (PGPR), and AM fungi can help in nutrient mobilization and confer tolerance to plants by alleviating adverse effects of stresses. Bradyrhizobia are mainly slow-growing, ubiquitous group of soil bacteria known as root symbiont; this is symbiotically associated with roots of soybean plants. In this chapter, we provided on how drought affects the soybean production worldwide and utilized specific bradyrhizobial strains to confer tolerance to soybean plants under drought stress and to understand the mechanisms imparting in the reduction of abiotic stress, e.g., GOGGAT MAPK, different polysaccharides, and other precursors involved in drought stress recovering mechanism. The enhancement of the soybean bradyrhizobial symbiosis participating in the drought tolerance particularly with climate smart bradyrhizobia having high osmotolerant traits, persisting longer in the field, and availability of such inoculants have also been discussed.

Keywords Bradyrhizobia · PGPR · AM fungi · Abiotic stress

A. Bharti · R. Agnihotri · H. S. Maheshwari · M. P. Sharma (✉)
ICAR-Indian Institute of Soybean Research, Indore, India

A. Prakash
Department of Microbiology, Faculty of Life Sciences, Barkatullah University, Bhopal, India

D. K. Choudhary et al. (eds.), *In Silico Approach for Sustainable Agriculture*,
https://doi.org/10.1007/978-981-13-0347-0_7

7.1 Introduction

Legumes have been used in farming since ancient times and have been among the first sources of human food. Soybean [*Glycine max* (L.) Merril] containing about 40% protein and 20% oil is emerging as one of the fastest-growing oilseed crops in the world. The worldwide soybean production in 2017–2018 is expected to be 348.04 million metric tons (Mt) (Global soybean production.com 2017). In India, Malwa plateau of Central India is the hub for soybean cultivation; the total production during 2014 increased to about 11.64 Mt from 10.02 Mt (Anonymous 2014–2015). However, the expansion of soybean in different parts of the country resulted in parallel growth of oil industry, for the past few years, soybean is facing climatic challenges, resulting in the decline of yield. When compared to other countries, the productivity of soybean per unit area is very low in India. It is stagnated mainly due to recurrence of drought, low nutrient use efficiency of crop, nutrient deficiency in soil, and abiotic stresses. Various studies have shown that drought stress leads to a decrease in soybean yield by up to 50% (Bagheri and Sadeghipour 2012). Plants respond differently at cellular, physiological, and biochemical levels to overcome the abiotic stresses. According to FAO's report, 2013, drought has become more frequent and intense globally and is a major factor causing alterations in plant at morphological, physiological, nutritional, and metabolic level, ultimately affecting the plant's overall growth and development adversely (Evelin et al. 2009). Drought stress in combination with heat stress causes enormous economic losses to farmers (Rizhsky et al. 2002; Das et al. 2016).

Soybean, a member of Fabaceae family, is able to form a symbiosis with bacteria belonging to the family *Rhizobiaceae*. The soybean inhabiting frequently with *Bradyrhizobium japonicum* is identified and documented as a slow-growing strain (Elkan and Bunn 1992). Based on differences in phenotypic characteristics, DNA sequences, production of rhizobiotoxines, composition of exopolysaccharides, presence of indole acetic acid, and other biochemical characteristics, soybean rhizobia (Young 1996; Sahgal and Johri 2003; Willems 2006) were categorized as follows:

Slow growers	*Bradyrhizobium japonicum*
	Bradyrhizobium elkanii
Extra-slow growers	*Bradyrhizobium liaoningense*
Fast growers	*Sinorhizobium fredii*
	Sinorhizobium xinjiangese
Moderate	*Mesorhizobium tianshanensis*

Bradyrhizobium spp. are slow growers, while *Sinorhizobium* spp. are fast growers (Kuykendall et al. 1992; Jordan 1982; Keyser et al. 1982). However, *Bradyrhizobium liaoningense* is recognized as extra-slow growing (Xu et al. 1995). The beneficial traits of rhizobia in soybean that have been studied in the recent past include root nodule formation and improved growth attributes (Saeki et al. 2006). However, selection of elite strains acclimatized to local environmental conditions and to newly bred plant lines stays a need of an hour (Appunu et al. 2008).

Among the factors controlling nitrogen fixation, drought alone is one of the most important factors influencing nodulation and survival in natural environments including soils. Globally, drought causes a major problem for nodulation of soybean which resulted in declined in production.

Recent developments have enhanced our understanding on root architecture and nodule traits in relation to impact of drought on soybean (Kunert et al. 2016). Nodule formation of rhizobial plants during the drought stress at physiological or molecular level is highly affected by and found to be extremely sensitive to drought stress. There is a need to understand mechanisms involved in the occurrence of drought-induced biological nitrogen fixation (BNF) inside the root nodules. It has been reported that drought stress causes reduced carbon assimilation due to stomata closure, membrane damage, and distressed activity of various CO_2-fixing enzymes, and heat stress increases membrane scratch and impairs metabolic functions (Farooq et al. 2009). Legumes, including soybean, have the ability to associate symbiotically with nitrogen-fixing bacteria needed for BNF which enables plants to take up atmospheric nitrogen and have agricultural, ecological, and environmental significance for sustaining the productivity of legumes. In this chapter we provided information on the potential role of *Bradyrhizobium* sp.-inoculated soybean on root nodulation and growth in response to drought (Clement et al. 2006) and understand the mechanisms involved while interacting with other plant growth-promoting microbes as part of microbiome influencing the tripartite symbiosis for enhancing drought tolerance in plants.

7.2 Drought as Limiting Factor in the Productivity of Soybean

Abiotic stresses include drought, winds, extreme temperatures, soil salinity, and flood. Among these stresses, drought is one of the most serious abiotic stresses which reduces crop productivity and quality (Vaishnav et al. 2017). According to Abate et al. (2012), most of commercial output comes from the export of soybean (83.8%), common bean (8.8%) groundnut (peanut) (4.9%), and chickpea (2.4%), but due to drought stress, it becomes obligatory due to inadequate rainfall or irrigation water and hence becomes one of the major factors limiting the crop production for realizing the food security. Water scarcity is a severe abiotic constraint of legume crops' productivity (Daryanto et al. 2015); therefore, developing drought-tolerant crops is a major research challenge both for plant scientists and agriculturists. Many reports have shown that drought leads to restriction of nodule formation at initial stages which imposes nodule senescence and nodule functions (Streeter 2003; Valentine et al. 2011) which later also affect leaves. Drought stress have detrimental effects on total biomass, pod number, seed number, seed weight and quality, and seed yield per plant. Drought stress 40% reduction in soybean's vegetable oil and protein feed yield (Valentine et al. 2011; Friedman and Brandon 2001) causing

enormous economic losses to farmers (Rizhsky et al. 2002). However, some varieties have capability to maintain leaf area, which provides survival and stability/capability to soybeans after the stress is relieved (Manavalan et al. 2009). They further suggested identifying traits associated with drought resistance in soybean where with the availability of the whole-genome sequence, physical maps, genetics and functional genomics tools, integrated approaches using molecular breeding, and genetic engineering offer new opportunities for improving drought resistance in soybean.

7.3 Role of Bradyrhizobia in the Alleviation of Drought Stress in Soybean

The non-judicious use of chemical fertilizers particularly in Indian agriculture resulted in negative effects on soil environments which have become a major concern for sustaining the crop and soil productivity. There is a need to look for an alternative strategy to sustain the productivity of oilseeds without disturbing the ecological balance. One such approach could be use of plant growth promoting microbes either signly or in combination of microorganisms, i.e., cocktail of two or more bacteria (consortia) performing multi-functions such as fixing nitrogen from atmospheric, solubilize or mobilize phosphorus, zinc and other soil nutrients to sustain plant growth and improve soil health (Babalola 2010; Sharma et al. 2010).

Since plants cannot use direct atmospheric nitrogen (N_2), they must use N_2 compounds such as nitrates. In order for organisms to use atmospheric N_2, it must be "fixed" or converted into ammonia (NH_3). This can happen occasionally through a lightning strike, but the bulk of nitrogen fixation is done by free-living or symbiotic bacteria. The nitrogenous enzyme present in these bacteria catalyzes the production of ammonia by combining gaseous nitrogen with ammonia. It is then further converted by the bacteria to make their own organic compounds. Some N_2-fixing bacteria live in the root nodules of legumes where they produce ammonia in exchange for sugars.

Since breeding of drought-tolerant lines is a long-term strategy, therefore the use of climate smart rhizobia capable of conferring tolerance to plants could be a viable solution. The integration of rhizobia with drought-tolerant lines would be more promising to cope with drought. Rhizobial inoculation is a common practice in agricultural legume production which requires survival and establishment of inoculated rhizobia in the soil environment (Catroux et al. 2001). *Soybean bradyrhizobia* are genus of gram-negative aerobic soil symbiotic bacteria associating with soybean which are mainly classified into two major groups: viz., the slow growing one which includes *Bradyrhizobium* spp. and the fast growing one which includes *Sinorhizobium* spp. (Sharma et al. 2010). *Bradyrhizobium japonicum* is a slow-growing most popular rhizobial strain, commercially exploited and widely used in soybean for improved root nodulation and N_2 fixation. In plants low molecular

weight osmolytes, including glycine betaine, proline, and other amino acids, organic acids, and polyols, play a vital role in sustaining cellular functions under drought. Plant growth substances such as salicylic acid, auxins, gibberellins, cytokinins, and abscisic acid modulate plant responses toward drought (Reibach and Streeter 1983) Polyamines, citrulline, and several enzymes act as antioxidants and reduce adverse effects of drought. Sucrose is the main carbohydrate observed in soybean nodules (Reibach and Streeter 1983), and other studies proved that sucrose accumulates in abiotic-stressed soybean nodules (Gonzalez et al. 1995, 1998). Some strains of *Bradyrhizobium japonicum* produce a different polysaccharide like rhamnose, galactose, and 2-O-methylglucuronic acid, inside the soybean root nodules (Streeter et al. 1992). Osmotic adjustments as a consequence of drought were evident from sevenfold increase in sugar level and showed similar increasing trend during stress which means sugars play a crucial role in drought stress; thus, a high sugar level may help in pea plant to survive under stress. Proline level increased 4–40 times as a consequence of drought, which concludes that this amino acid may help plants to survive during stress conditions (Sanchez et al. 1998). The effect of drought stress on the nodule inhabiting bradyrhizobial strain development and stability or BNF remains unclear, particularly at the molecular level. In soybean, drought stress also results in the accumulation of trehalose (Streeter 2003); trehalose is a disaccharide synthesized by many bacteria like rhizobia which significantly participate during stressed periods to stabilize membrane and protein structure during desiccation. Bradyrhizobial species, e.g., *Bradyrhizobium japonicum* and *B. elkanii*, harbor three enzymes needed for trehalose synthesis as trehalose synthase (TS), maltooligosyl trehalose synthase (MOTS), and trehalose-6-phosphate synthetase; from these enzymes, TS was found to be the dominant enzyme in bradyrhizobia where maltose concentration is very low which results in higher accumulation of trehalose and hence very important to cope up with drought (Streeter and Gomez 2006).

7.4 Co-inoculation of Soybean Bradyrhizobia with Other Plant Growth-Promoting Microbes

7.4.1 Interaction with PGPR (Plant Growth-Promoting Rhizobacteria)

The term PGPR was coined by Kloepper and Schroth (1981), where all PGPR can manage major activities in plants necessary for their survival in nature, e.g., can help in plant growth and mineral nutrition and increase resistance against stresses. Numerous reports are available where PGPR have been applied in agriculture, resulting in increased seedling emergence, plant weight, crop yield, and disease resistance (Kloepper et al. 1999). PGPR help plants in many ways through stimulation of PGP traits and modulate osmotic adjustments in the growing plants to

overcome the effects of drought stress. The success of the soybean crop in Brazil is one such example where *Bradyrhizobium* is attributed to be the most symbiotic benefit imparting in the BNF process and fulfilling the requirement of nitrogen for ensuring high yields without supplementing mineral nitrogen fertilizers (Alves et al. 2003; Hungria et al. 2006).

Inoculation of PGPR enhances root and plant growth and nutrient content through plant's root architecture. The root portion of the plant has many contributions; e.g., root hairs help in uptake of nutrients (Lauter et al. 1996); in Fabaceae the root tip starts the rhizobial colonization which sooner or later leads to formation of a root nodule (Desbrosses and Stougaard 2011). Root system architecture (RSA) together with PGPR strains like *Pseudomonas* or *Bacillus* helps the plant to grow faster under unfavorable circumstances and provide plant fitness against stresses. The combined inoculation of PGPR strain with arbuscular mycorrhizal fungi (AM fungi) (*Pseudomonas putida* and *Gigaspora rosea*) particularly having 1-aminocyclopropane-1-carboxylate (ACC) deaminase-positive trait showed enhanced plant growth and developed plant root architecture. The results also showed that ACC deaminase-producing PGPR which isolates along with AM fungi can enhance survival and adaptability of plants under stress. PGPR inhabit plant root nodule and influence plant productivity and immunity; however, recent work by several groups showed that PGPR also enhances systemic tolerance to drought (Yang et al. 2009). Specific combinations of autochthonous or allochthonous inoculants also contribute to plant drought tolerance by changing proline and antioxidative activities. However, un-inoculated plants have low relative water and nutrient content, shoot proline accumulation, and glutathione reductase activity, but the higher superoxide dismutase activity, stomatal conductance and electrolyte leakage, and microbial activities irrespective of the microbial origin seem to be coordinated with the functioning of the plant as an adaptive response through modulating water stress tolerance and minimizing the cellular damages (Ortiz et al. 2015).

7.4.2 Interaction with AM Fungi

The efficiency of the tripartite symbiosis formed by soybean AM fungi and *Rhizobium* needs to be critically examined to determine the most effective combination for maximizing plant growth. Betiana et al. (2015) suggested that co-inoculation of different AM fungal strains (*Septoglomus constrictum*, *Glomus* sp., and *Glomus aggregatum*) did not performed effectively to cope with drought stress in soybean but rather found almost equal or better results when inoculated individually. However some studies have also shown contradictory results where a mixture/consortium of different AM fungal species produced better effects in comparison to single inoculations (Hoeksema et al. 2010; Verbruggen and Kiers 2010). Plants benefit with necessary nitrogen by establishing symbiotic relationship with *Rhizobium*. On the other hand, AM fungal symbiosis contributes to plant growth promotion by hyphal-mediated nutrient and water uptake from soil (Smith and Read 2008).

Interaction between AM fungi and rhizobia benefits the legumes in terms of nitrogen fixation as well as P uptake (Bethlenfalvay and Yoder 1981; Bethlenfalvay et al. 1982). *Bradyrhizobium japonicum* has been found to be synergistic to AM fungi by causing stimulatory effect on *Glomus mosseae* (GM) colonization via Nod factor production leading to accretion of flavonoid, where equally positive effects were derived upon exogenous addition of Nod factors and flavonoid (Xie et al. 1995). Inoculation of AM fungi has been shown to result in betterment of droughted soybean plants conferring drought tolerance as MDA content; i.e., a measure of lipid peroxidation was found to be lower in nodules (Ruiz-Lozano et al. 2001), root (Porcel et al. 2003), and shoot (Porcel and Ruiz-Lozano 2004) as a consequence of AM fungi inoculation.

Meghvansi et al. (2008) found the combination of *Glomus intraradices* and *Bradyrhizobium japonicum* is most effective in increasing nodulation as well as productivity where both sole and dual inoculations were equally efficient in increasing shoot N, shoot height, and biomass. Babalola et al. (2009) noted that in improving growth attributes of soybean such as shoot N and P and nodule biomass as well as soil nutritional profile and organic matter content, GM and *Bradyrhizobium japonicum* formed a more efficient pair than *Glomus deserticola* and *Bradyrhizobium japonicum.* The combination of *Rhizobium* and *Glomus fasiculatus* has been shown to improve nodulation, shoot biomass, and N_2 content in soybean (Bagyaraj et al. 1979). However in all the above studies, the symbiotic efficiency of these two wasn't tested under stressed conditions.

7.4.3 AM Interaction Under Drought Stress

Under drought conditions, Ruiz-Lozano et al. (2001) found that AM alone or in combination with *Bradyrhizobium* proved a better colonizer for soybean where the effects were evident from higher arbuscule richness. It was evident from the studies of Ivanov et al. (2012) that the symbiotic relationship between *Rhizobium* and leguminous plant could play a role in a pathway that is crucial to arbuscule formation. Shoot biomass which depicts plant growth has been used as an indicator to assess drought-induced effects. Under drought stress, in terms of growth, co-inoculation of AM fungal species *Glomus mosseae* (GM) with *Bradyrhizobium japonicum* did better and brought about 15% increase in shoot biomass than corresponding singly inoculated plants (Ruiz-Lozano et al. 2001). However, in the same study, *Glomus intraradices* increased shoot biomass and N_2 content, when inoculated in combination with *Bradyrhizobium* as compared to its sole inoculation. Drought-stressed soybean plants when co-inoculated with *Bradyrhizobium japonicum* and GM possessed higher nitrogenase activity evident from acetylene reduction assay (ARA) as compared to the plant having single inoculation with *Bradyrhizobium japonicum* (Porcel et al. 2003). Ding et al. (2012) studied co-inoculation of soybean plants with GM, and *Bradyrhizobium japonicum* increased plant phosphorus content which in turn maintained a positive correlation

with proton release from nodules and hyphae, and it was suggested that increased nitrogenase activity brought about by AM fungi inoculation leads to increased proton discharge.

Under stress conditions, plant undergoes several adverse conditions that could hamper this symbiotic efficiency, and adverse effects could be experienced in the form of nodulation, photosynthesis, and nitrogen fixation in soybean (Sprent 1971). The factors that affect the symbiotic efficiency can include soil type. Soil nutritional profile should be examined prior to inoculation. Soil nutrient deficiency supports AM fungi growth (Lovelock et al. 2003) as soils with elevated P levels reduce mycorrhizal colonization (Smith and Read 2008). Rhizobium-AM fungal symbiosis could promote the survival of soybean plant in soil with lower nutrient profile particularly phosphorus-/nitrogen-deficient ones by increasing shoot biomass and plant nitrogen and phosphorus content (Wang et al. 2011). However the effects were not evident when soils had sufficient N and P levels.

AM fungi species alone or in combination might not perform adequately depending on one or the other reasons mentioned above. The varying patterning of growth between different species of AM fungi could also lead to differential growth effects on plants as well as the symbiotic efficiency of other microbial partner upon dual inoculation. In the experiments of Ruiz-Lozano et al. (2001) under drought-stressed as well as unstressed conditions, soybean plants co-inoculated with *Glomus intraradices* and *Bradyrhizobium japonicum* were lagging behind in terms of nodule growth, total lipid content, and arbuscule richness. Because *Glomus intraradices* invests more nonstructural carbohydrates in the development of intra-radical structures, when inoculated with *Bradyrhizobium*, as a consequence of which, a competition was generated between the two microsymbionts (Ruiz-Lozano et al. 2001). There have been reports of competition for photosynthates between *Rhizobium* and *Glomus* in soybean (Harris et al. 1985). The inoculation strategy also matters when it comes to deriving maximum benefits of AM fungi-*Bradyrhizobium* symbiosis. Bethlenfalvay et al. (1985) demonstrated that there might exist an antagonistic effect in symbiosis establishment between *Glomus* and *Bradyrhizobium* when either member of the pair had colonized the root before the other. Ballesteros-Almanza et al. (2010) showed that independent inoculation of *Glomus intraradices* and *Rhizobium* could prove a better strategy for achieving drought tolerance in some common beans which was evident from decrease in biomass resulting from co-inoculation under the conditions of drought. However AM fungi colonization is positively correlated with trehalose content in nodules.

Plants are classified into two categories based on their ability to resist drought, i.e., drought avoiders that cope up with conditions of water deficit either via increasing water uptake or by minimizing water deficit owing to low dehydration tolerance limit (maintenance of high internal water potential) and drought tolerators that have low drought avoidance property and where internal water potential is considerably low (Ludlow 1989). Drought avoidance is the chief mechanism via which mycorrhizal plants alleviate drought stress, and AM fungi-inoculated plants are shielded against severe effects of drought stress as a consequence of hyphal-mediated water uptake as studied in drought-exposed soybean plants (Porcel et al.

2003). Mycorrhizal plants are in a "hydrated state" acquired via water absorption and subsequent transport to plant by AM fungi hyphae (Ruiz-Lozano and Azcon 1995). Aliasgharzad et al. (2006) pointed out that AM fungi-inoculated plants use drought avoidance to mitigate stress caused by water deficit in plants, and this mechanism could come into play, even in plants dually inoculated with *G. etunicatum* and *B. japonicum*. In their study it was observed that soybean plants having dual inoculation of these two microbes had higher relative water content and leaf water potential which are the strategies involved in drought avoidance mechanism, and this resulted in promotion of growth attributes (i.e., higher shoot and seed biomass).

Cellular components get severely affected by reactive oxygen species (ROS) such as superoxide, hydrogen peroxide, and hydroxyl radicals leading to oxidative stress (Apel and Hirt 2004). Antioxidant enzymes form an important defense mechanism against stress. Under drought stress, mycorrhizal lettuce plants had higher superoxide dismutase activity (Ruiz-Lozano et al. 1996). Mediated resistance to oxidative stress was also studied by Bressano et al. (2010) where AM fungi-inoculated soybean plants were protected against detrimental effects, i.e., oxidative stress caused by herbicide paraquat (PQ), as none of the growth parameters were affected by stress, and root and shoot length were significantly increased. In this study root growth is correlated with mycorrhizal colonization. AM fungi-inoculated plants also had lower leaf malondialdehyde content (MDA) which gives an estimate of lipid peroxidation. However Aliasgharzad et al. (2006) found that GM was found to be less efficient than *Glomus etunicatum* in maintaining higher leaf relative water content under moisture deficit stress.

7.5 Mechanism Involved

Soybean is mainly grown under rain-fed conditions, and due to erratic and uncertainty in rainfall pattern, soybean has to face frequent droughts, and due to its sensitivity to drought stress, its yield and seed quality reduce significantly (Purcell et al. 2004). However, during abiotic stress tolerance, plants apply different adaptive strategies to adjust the effect of inadequate water supply (Kramer and Boyer 1995). Uma et al. (2013) reported that some of *Bradyrhizobium* isolates screened at 30% PEG concentration and nodule drought tolerance has been linked to the ability to sustain a supply of photosynthate to the nodules during drought stress mechanism (King and Purcell 2001). A number of studies on root development where a greater root to shoot ratio obtained as soil dries indicating enhanced root growth (Sharp et al. 2004). Similarly, due to low water potential, root parts become thinner (Sharp et al. 1988), and the process becomes complicated where root hairs begin to turn over and enveloped some of the bacteria. The mechanism of BNF inside the nodule is based on the carbon contribution via sucrose synthase and the O_2 availability via leghemoglobin (Lb) and the nitrogen (N_2) feedback regulation (Serraj et al. 1999). Glutamine synthetase (GS) is the precursor of N_2 assimilation in plants, and GS1 only participated in the symbiotic assimilation of ammonia (Morey et al. 2002).

Nitrogenase enzyme promotes conversion of N_2 to ammonium (NH_4), and this activity depends on ATP (adenosine-5′-triphosphate) and glutamate synthase (GOGAT). Although in plants excess amount of amino acids showing assimilation of nitrogen in form of ammonium, along with higher activity of GS and GOGAT enzymes (Ramos et al. 2005).

The bradyrhizobia are host specific; they inhabit soybeans and are regulated by some identified compounds that are released by soybean roots and are fascinated to those compounds (Spaink 1994). Once they're in close proximity with the roots, the bacteria will, in turn, begin releasing other compounds. Bradyrhizobia exhibited alleviation of moisture stress in plants, and bradyrhizobia impart drought tolerance by producing exopolysaccharides (EPS), phytohormones, 1-aminocyclopropane-1-carboxylate (ACC) deaminase and volatile compounds, inducing accumulation of osmolytes and antioxidants, upregulation or downregulation of stress-responsive genes, and alteration in root morphology in acquisition of drought tolerance; with this, plant's secreted phytohormones are very important for the plants' growth and development, such as IAA, gibberellins, ethylene, abscisic acid (ABA), and cytokinins. The term induced systemic tolerance (IST) was coined for physical and chemical changes induced by microorganisms in plants which results in enhanced adaptability against drought stresses (Yang et al. 2009; Vurukonda et al. 2016). The signaling pathway of drought stress is mainly intermingled with the signaling pathway of osmotic stresses (Ahuja et al. 2010). ABA is produced by various parts of the plant from the root, parenchyma, and mesophyll under drought stress (Wilkinson and Davies 2010) where ABA regulates the closing of stomata, and the plant is able to cope up with environmental stresses (Zhang et al. 2006).

The experiment of Brechenmacher et al. (2010) concluded that nodulation is a complex process in soybean (*Glycine max*) root; this bacteroid *Bradyrhizobium japonicum* inside the root mutually coordinated by signal molecule analyzed by the different chromatographic technique characterized 2610 metabolites in root hairs like flavonoids, amino acids, fatty acids, carboxylic acids, and various carbohydrates. And trehalose was among the most strongly induced metabolites produced following inoculation. Due to osmotic or desiccation stress, bacteria build up osmoprotective compounds referred to as osmolytes. Osmolytes can be uptaken from the environment (exogenous) or through de novo biosynthesis (endogenous). De novo biosynthesis is a common response to desiccation stress in many bacteria (Wood et al. 2001). McIntyre et al. (2007) investigated the role of endogenous trehalose synthesis in desiccation tolerance in *R. leguminosarum* bv. *trifolii* strain NZP561. Strain NZP561 accumulated trehalose as it entered the stationary phase due to the combined actions of the Tre YZ and Ots AB pathways. Although importance of various trehalose biosynthetic pathways and their contribution in stress tolerance of the bacteria have not been identified, accumulated trehalose plays an important role in protecting *R. leguminosarum* bv. *trifolii* cells against desiccation stress (Fig. 7.1).

Many researchers reported importance of protein kinases in drought stress mechanism (Bartels et al. 2010). Mitogen-activated protein kinase (MAPK) cascades play important roles in the stress management in both plants and microorganisms. The

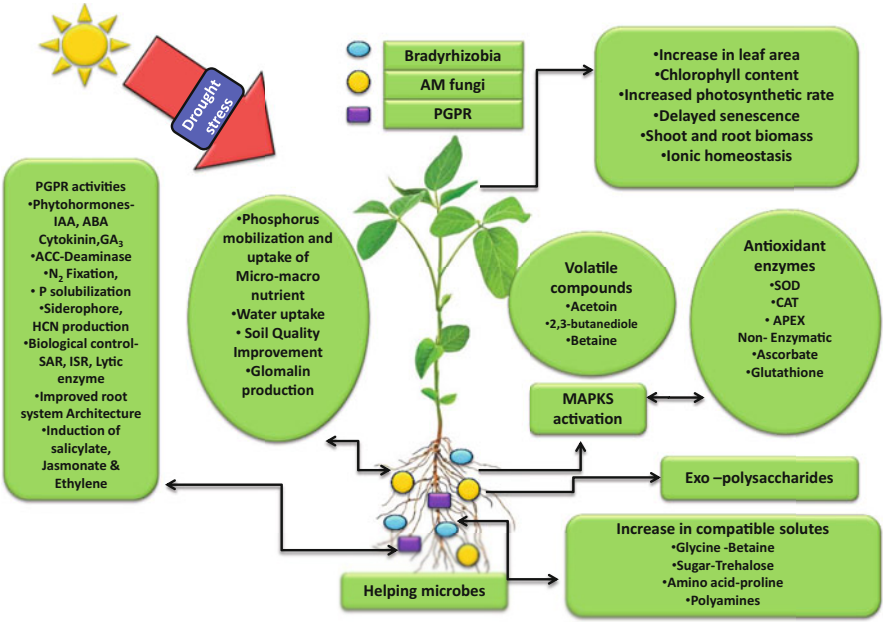


Fig. 7.1 Notified different supporting mechanism involved in drought stress management

bradyrhizobial symbiosis is established with the plants which can enhance plant drought tolerance and might be closely related to the microbial MAPK response and the molecular dialogue between bacteria and soybean MAPK cascades. Many studies support the role of protein kinases in stress signaling (Bartels et al. 2010). In plants, drought stress-responsible signals of the MAPK family (MAPK, MAPKK/MEKK, MAPKKK/MKK) as well as the MAPK phosphatases have been well studied in *Arabidopsis thaliana* (Boudsocq and Lauriere 2005) even though a PA-responsible MAPK has been recognized in soybean but still requires more studies (Lee et al. 2001). Morphological or physiological management is sometimes applied to avoid drought stress, and different associated traits of roots have been understood as indicators of drought tolerance in soybean (Garay and Wilhelm 1983).

Enhanced glutathione reductase activity in nodules of GM and *Bradyrhizobium japonicum*-inoculated soybean plants under drought stress was suggested as a mechanism to combat premature nodule senescence in soybean plants facing drought stress (Porcel et al. 2003). However authors did suggest the possibility of existence of other possible mechanisms as well. One such mechanism could be the ability of mycorrhizal soils to retain water as compared to corresponding non-mycorrhizal counterparts (Augé et al. 2001a, b).The property of soils to retain water as a consequence AM fungi inoculation is due to enhanced soil aggregation achieved via increased macro-aggregation by AM fungi hyphae (Tisdall and Oades 1982) and also due to glomalin production (Wright 2000). Glomalin is a soil protein produced on hyphal walls of AM fungi (Driver et al. 2005), where its production has

been found to increase under stressed conditions (Gadkar and Rillig 2006), and as the protein is thought to be produced by fungus to modulate the microenvironment to boost its own growth (Rillig and Steinberg 2002), it could also result in a better aggregated soil thus protecting the plants against water-deficit stress. Glomalin contains significant amount of carbon as well as nitrogen (Lovelock et al. 2004) which indicates the possibility of its increased production in the presence of nitrogen-fixing bacterium like *Bradyrhizobium japonicum*. Another strategy to mitigate drought stress is through osmotic adjustment. Osmotic adjustment is achieved via solutes such as K^+, proline, glycine betaine, and carbohydrates like sucrose, pinitol, mannitol, etc. (Aliasgharzad et al. 2005; Ruiz-Lozano 2003). In order to protect soybean plant from oxidative stress, AM fungi inoculation should improve root water status, mainly via osmotic adjustment, and consequently maintain higher leaf water potential (Porcel and Ruiz-Lozano 2004). Aliasgharzad et al. (2006) found that soybean plants co-inoculated with *G. etunicatum* and *B. japonicum* possessed higher shoot potassium and nitrogen content, both of which are crucial to osmotic adjustment and stomatal conductance. There have been reports that potassium ion plays important role in both of these phenomena, i.e., osmotic adjustment and stomatal conductance (Ruiz-Lozano 2003). As observed in case of soybean, AM fungi and *Bradyrhizobium* when inoculated together resulted in improved P uptake and better soil quality (Tilak et al. 1995). Synergistic effect of co-inoculated bacteria and AM fungi helps in restoring plant growth under drought stress (Marulanda et al. 2009). Recent studies have corroborated a positive effect of the interactions between AM fungi and *Bradyrhizobia* under drought conditions (Table 7.1).

7.6 Conclusion and Future Research Directions

Drought is considered to be a major threat to soybean cultivation not only in India but also worldwide. Being a legume crop fulfill its nitrogen demand and attain overall growth through biological nitrogen fixation performed through root-nodulating soybean rhizobia. During drought conditions, the roots' development is affected adversely which consequently affects the soybean-rhizobial symbiosis, root nodulation, and ultimately overall productivity (Kunert et al. 2016). The long-term strategy of breeding drought-tolerant lines encouraged to look for a quick solution where microbial strategy involving robust soybean rhizobia could be a potential and short-term solution to some extent to cope with drought and sustain soybean productivity. The role of customized and climate smart soybean rhizobia capable of surviving under low-moisture conditions and improving the overall soybean-rhizobia symbiosis is very important and the need of an hour. In this chapter we aimed to provide information on how drought imparts changes in plant physiology and indicators which can be assessed to recover robust rhizobia with better osmotic tolerance traits to persist longer in the harsh drought environment and under low-moisture stress conditions. Further the interaction and co-inoculation effect of

Table 7.1 Example and role of *Bradyrhizobia* in conferring drought tolerance in soybean (*Glycine max* (L.) Merr)

Microbes	Inference	References
Bradyrhizobium japonicum	Effect of 4–8-week drought stress did not reduced population of *B. japonicum* but would affect soybean growth and development	Delaux et al. (2014)
Rhizobia spp.	Help *Rhizobia* reduced drought stress by changing root morphology, transpiration rate, and leaf stomatal conductance	Chi et al. (2005)
Bradyrhizobium japonicum, *B. elkanii*, and *B. diazoefficiens*	The consortia application fulfills the whole nitrogen requirement to sustain high yields	Hungria et al. (2006)
Bradyrhizobia strains	N_2 fixation at high temperatures	Rahmani et al. (2009)
Rhizobia spp.	Improved water holding capacity of the rhizosphere region to improve nutrient uptake and water flow	Sandhya et al. (2009)
Bradyrhizobium japonicum	Performed better in drought tolerance with basis of IAA, EPS production, nodulation, nodule ARA, nodule N_2 content	Uma et al. (2013)
Glomus intraradices and *Bradyrhizobium japonicum*	Improved growth, nodulation, soil phosphorus, productivity	Meghvansi and Mahna (2009)
Glomus intraradices	Higher root proline content, increased shoot biomass, lower lipid peroxidation in shoot	Porcel and Ruiz-Lozano (2004)
Endogone mosseae	Improved root water status via promotion of water via root, improved nutritional status, increased leaf relative water content and turgor	Safir et al. (1972)
Glomus intraradices, *Glomus mosseae*, and *Glomus caledonium*	Improved crop performance under drought conditions	Ortiz et al. (2015)

soybean rhizobia with other plant growth-promoting microorganisms (PGPMs), e.g., plant growth-promoting rhizobacteria and AM fungi, was dealt with examples to enhance drought tolerance and improve soybean productivity. The underlying mechanisms conferring the stress tolerance of rhizobial- and PGPM-mediated soybeans are also provided. Current developments have been prepared for our understanding of the drought effect on soybean plant and nodule, as well as with a view to improve the selection of more drought-tolerant soybean cultivars and *Bradyrhizobia* in the future. The soybean rhizobia accumulating higher-specific osmolytes such as higher accumulation of carbohydrates, e.g., trehalose, and selecting them under low-moisture stress conditions simulated under in vitro could be one of the potential robust soybean rhizobia capable of performing under drought conditions. Therefore, such strains and its integration with other compatible microbes to evolve as effective microbial consortia need to be evaluated under field eventually to be released as a robust rhizobia and microbial consortia for large-scale application in soybean.

Acknowledgments The authors would like to thank the Director of ICAR-Indian Institute of Soybean Research, Indore, for providing the necessary facilities. Funding from ICAR-AMAAS network subproject to MPS and fellowship to AB is gratefully acknowledged.

References

Abate T, Alene AD, Bergvinson D, Shiferaw B, Silim S, Orr A, Asfaw S (2012) Tropical grain legumes in Africa and South Asia: knowledge and opportunities. International Crops Research Institute for the SemiArid Tropics, Nairobi. 112 pp. ISBN: 978-92-9066-544-1

Ahuja I, De Vos RCH, Bones AM, Hall RD (2010) Plant molecular stress responses face climate change. Trends Plant Sci 15(12):664–674

Aliasgharzad N, Barin M, Samadi A (2005) Effects of NaCl–induced and mixture salinities on leaf proline and growth of tomato in symbiosis with arbuscular mycorrhizal fungi. In: Proceeding of the international conference on environmental management, Hyderabad, India. pp 203–207

Aliasgharzad N, Neyshabouri M, Salimi G (2006) Effects of arbuscular mycorrhizal fungi and Bradyrhizobium japonicum on drought stress of soybean. Biologia (Bratisl) 61:324–328. https://doi.org/10.2478/s11756-006-0182-x

Alves BJR, Boddey RM, Urquiaga S (2003) The success of BNF in soybean in Brazil. Plant Soil 252:1–9

Anonymous (2014–2015) Director's report and summary tables of experiments 2014–15, all India Coordinated Research Project on Soybean, ICAR–Directorate of Soybean Research, Indore India. p 329

Apel K, Hirt H (2004) Reactive oxygen species: metabolism, oxidative stress, and signal transduction. Annu Rev Plant Biol 55:373–399

Appunu C, N'Zoue A, Laguerre G (2008) Genetic diversity of native bradyrhizobia isolated from soybeans (*Glycine max* L.) in different agricultural–ecological–climatic regions of India. Appl Environ Microbiol 74:5991–5996

Augé RM, Kubikova E, Moore JL (2001a) Foliar dehydration tolerance of mycorrhizal cowpea, soybean and bush bean. New Phytol 151:535–541

Augé RM, Stodola AJW, Tims JE, Saxton AM (2001b) Moisture retention properties of a mycorrhizal soil. Plant Soil 230:87–97

Babalola OO (2010) Beneficial bacteria of agricultural importance. Biotechnol Lett 32:1559–1570

Babalola OA, Atayese MO, Soyoye T (2009) Influence of Bradyrhizobium and two Glomus species on the growth and yield of soybean. J Agric Environ Sci 9:79–95

Bagheri A, Sadeghipour O (2012) Biochemical changes of common bean (*Phaseolus vulgaris* L.) to pretreatment with salicylic acid (SA) under water stress conditions. Int J Biosci 2:14–22

Bagyaraj DJ, Manjunath A, Patil RB (1979) Interactions between a VAM and Rhizobium and their effects on soybean in the field. New Phytol 82:141–145

Ballesteros-Almanza L, Altamirano-Hernandez J, Peña-Cabriales JJ, Santoyo G, Sanchez-Yañez JM, Valencia-Cantero E, Macias-Rodriguez L, Lopez-Bucio J, Cardenas-Navarro R, Farias-Rodriguez R (2010) Effect of co–inoculation with mycorrhiza and rhizobia on the nodule trehalose content of different bean genotypes. Open Microbiol J 4:83–92

Bartels S, Besteiro MAG, Lang D, Ulm R (2010) Emerging functions for plant MAP kinase phosphatases. Trends Plant Sci 15(6):322–329

Bethlenfalvay G, Yoder J (1981) The Glycine–Glomus–Rhizobium symbiosis. Physiol Plant 52:141–145

Bethlenfalvay G, Pacovsky R, Bayne H, Stafford A (1982) Interactions between nitrogen fixation, mycorrhizal colonization, and host–plant growth in the Phaseolus–Rhizobium–Glomus symbiosis. Plant Physiol 70:446–450

Bethlenfalvay GJ, Brown MS, Stafford AE (1985) Glycine–Glomus–Rhizobium symbiosis. 11. Antagonistic effects between mycorrhizal colonization and nodulation. Plant Physiol 79:1054–1058

Betiana C, Grümberg, Urcelay C, María A, Shroeder, Vargas-Gil S, Luna CM (2015) The role of inoculum identity in drought stress mitigation by arbuscular mycorrhizal fungi in soybean. Biol Fertil Soils 51:1–10

Boudsocq M, Lauriere C (2005) Osmotic signaling in plants. Multiple pathways mediated by emerging kinase families. Plant Physiol 138(3):1185–1194

Brechenmacher L, Lei Z, Libault M, Findley S, Sugawara M, Sadowsky MJ, Sumner LW, Stacey G (2010) Soybean metabolites regulated in root hairs in response to the symbiotic bacterium Bradyrhizobium japonicum. Plant Physiol 153:1808–1822

Bressano M, Curetti M, Giachero L, Vargas GS, Cabello M, March G, Ducasse DA, Luna CM (2010) Mycorrhizal fungi symbiosis as a strategy against oxidative stress in soybean plants. J Plant Physiol 167:1622–1626

Catroux G, Hartmann A, Revellin C (2001) Trends in rhizobial inoculant production and use. Plant Soil 230:21–30

Chi F, Shen S, Cheng H, Jing Y, Yanni YG, Dazzo FB (2005) Ascending migration of endophytic rhizobia, from roots to leaves, inside rice plants and assessment of benefits to rice growth physiology. Appl Environ Microbiol 71:7271–7278

Clement M, Boncompagni E, Almeida-Engler JD, Herouart D (2006) Isolation of a novel nodulin: a molecular marker of osmotic stress in *Glycine max Bradyrhizobium japonicum* nodule. Plant Cell Environ 29:1841–1852

Daryanto S, Wang L, Jacinthe PA (2015) Global synthesis of drought effects on food legume production. PLoS One 10(6):1–16

Das A, Eldakak M, Paudel B, Kim DW, Hemmati H, Basu C, Rohila SJ (2016) Leaf proteome analysis reveals prospective drought and heat stress response mechanisms in soybean. Biol Med Res Int 6021047:1–23

Delaux CB, Marburger D, Delaux PM, Conley S, Ané JM (2014) Effect of drought on Bradyrhizobium japonicum populations in Midwest soils. Plant Soil 382:165–173

Desbrosses GJ, Stougaard J (2011) Root nodulation: a paradigm for how plant–microbe symbiosis influences host development al path–ways. Cell Host Microbe 10:348–358

Ding X, Sui X, Wang F, Gao J, He X, Zhang F, Yang J, Feng G (2012) Synergistic interactions between Glomus mosseae and Bradyrhizobium japonicum in enhancing proton release from nodules and hyphae. Mycorrhiza 22:51–58. https://doi.org/10.1007/s00572-011-0381-3

Driver JD, Holben W, Rillig MC (2005) Characterization of glomalin as a hyphal wall component of arbuscular mycorrhizal fungi. Soil Biol Biochem 37:101–106

Elkan GH, Bunn CR (1992) The rhizobia. In: Balows A, Truper HG, Dworkin M, Harder W, Schleifer K-H (eds) The prokaryotes. A handbook on the biology of bacteria: ecophysiology, isolation, identification, applications, 2nd edn. Springer–Verlag, New York, pp 2197–2213

Evelin H, Kapoor R, Giri B (2009) Arbuscular mycorrhizal fungi in alleviation of salt stress: a review. Ann Bot 104:1263

Farooq M, Wahid MA, Kobayashi N, Fujita D, Basra SMA (2009) Plant drought stress: effects, mechanisms and management. Agron Sustain Dev 29:185–212

Friedman M, Brandon DL (2001) Nutritional and health benefits of soy proteins. J Agric Food Chem 49:1069–1086

Gadkar V, Rillig MC (2006) The arbuscular mycorrhizal fungal protein glomalin is a putative homolog of heat shock protein 60. FEMS Microbiol Lett 263:93–101. https://doi.org/10.1111/j.1574–6968.2006.00412.x

Garay AF, Wilhelm W (1983) Root system characteristics of two soybean isolines undergoing water stress condition. Agron J 75:973–977

Global Soybean Production: The Global Soybean production (2017) http://www.globalsoybeanproduction.com. Accessed 19 Jan 2017

Gonzalez EM, Gordan AJ, James CL, Igor CA (1995) The role of sucrose synthase in the response of soybean nodule to drought. J Exp Bot 46:1515–1524

Gonzalez EM, Aparicio-Tejo PM, Gordon AJ, Minchin FR, Royuela M, Igor CA (1998) Water–deficit effect on carbon and nitrogen metabolism of pea nodule. J Exp Bot 49:1705–1714

Harris D, Pacovsky RS, Paul EA (1985) Carbon economy of soybean– Rhizobium–Glomus associations. New Phytol 101:427440

Hoeksema JD, Chaudhary VB, Gehring CA, Johnson NC, Karst J, Koide RT, Pringle A, Zabinski C, Bever JD, Moore JC, Wilson GWT, Klironomos JN, Umbanhowar J (2010) A meta–analysis of context dependency in plant response to inoculation with mycorrhizal fungi. Ecol Lett 13:394–407

Hungria M, Campo RJ, Mendes IC, Graham PH (2006) Contribution of biological nitrogen fixation to the N nutrition of grain crops in the tropics: the success of soybean (*Glycine max* L. Merr.) in South America. In: Singh RP, Shankar N, Jaiwa PK (eds) Nitrogen nutrition and sustainable plant productivity. Studium Press, Houston, pp 43–93

Ivanov S, Fedorova EE, Limpens E, De Mita S, Genre A, Bonfante P, Bisseling T (2012) Rhizobium–legume symbiosis shares an exocytotic pathway required for arbuscule formation. Proc Natl Acad Sci 109:8316–8321. https://doi.org/10.1073/pnas.1200407109

Jordan DC (1982) Transfer of Rhizobium japonicum Buchanan 1980 to Bradyrhizobium gen. nov., a genus of slow–growing, root nodule bacteria from leguminous plants. Int J Syst Bacteriol 32:136–139

Keyser HH, Bohlool BB, Hu TS, Weber DF (1982) Fast-growing rhizobia isolated from root nodules of soybean. Science 215:1631–1632

King CA, Purcell LC (2001) Soybean nodule size and relationship to nitrogen fixation response to water deficit. Crop Sci 41:1099–1107

Kloepper JW, Schroth MN (1981) Plant growth promoting rhizobacteria and plant growth under gnotobiotic condition. Phytopathology 71:642–644

Kloepper JW, Rodriguez-Kabana R, Zehnder GW, Murphy J, Sikora, Fernandez C (1999) Plant root–bacterial interactions in biological control of soilborne diseases and potential extension to systemic and foliar diseases. Australas Plant Pathol 28:27–33

Kramer PJ, Boyer JS (1995) Water relations of plant and soils. Academic, New York

Kunert KJ, Vorster BJ, Fenta BA, Kibido T, Dionisio G, Foyer CH (2016) Drought stress responses in soybean roots and nodules published. Front Plant Sci 7(1015):1–7

Kuykendall LD, Saxena B, Cevine TE, Udell SE (1992) Genetic diversity in Bradyrhizobium. Jordan 1982 and a proposal for *Bradyrhizobium elkanii* sp. nov. Can J Microbiol 38:201–505

Lauter FR, Ninnemann O, Bucher M, Riesmeier JW, Frommer WB (1996) Preferential expression of an ammonium transporter and of two putative nitrate transporters in root hairs of tomato. Proc Natl Acad Sci U S A 93(15):8139–8144

Lee S, Hirt H, Lee Y (2001) Phosphatidic acid activates a wound-activated MAPK in *Glycine max*. Plant J 26(5):479–486

Lovelock CE, Andersen K, Morton JB (2003) Arbuscular mycorrhizal communities in tropical forests are affected by host tree species and environment. Oecologia 135:268–279

Lovelock CE, Wright SF, Clark DA, Ruess RW (2004) Soil stocks of glomalin produced by arbuscular mycorrhizal fungi across a tropical rain forest landscape. J Ecol 92:278–287

Ludlow MM (1989) Strategies in response to water stress. In: Kreeb HK, Richter H, Hinkley TM (eds) Structural and functional response to environmental stresses, water shortage. SPB Academic Press, Netherlands, pp 269–281

Manavalan LP, Guttikonda SK, Tran LSP, Nguyen HT (2009) Physiological and molecular approaches to improve drought resistance in soybean. Plant Cell Physiol 50(7):1260–1276

Marulanda A, Barea JM, Azcon R (2009) Stimulation of plant growth and drought tolerance by native microorganisms (AMF and bacteria) from dry environments: mechanisms related to bacterial effectiveness. J Plant Growth Regul 28:115–124

McIntyre HJ, Davies H, Hore HA, Miller SH, Dufour JP, Ronson CW (2007) Trehalose biosynthesis in rhizobium leguminosarum bv. trifolii and its role in desiccation tolerance. Appl Environ Microbiol 73(12):3984–3992

Meghvansi MK, Mahna SK (2009) Evaluating the symbiotic potential of Glomus intraradices and Bradyrhizobium japonicum in vertisol with two soybean cultivars. Am Eurasian J Agron 2:21–25

Meghvansi MK, Prasad K, Harwani D, Mahna SK (2008) Response of soybean cultivars toward inoculation with three arbuscular mycorrhizal fungi and Bradyrhizobium japonicum in the alluvial soil. Eur J Soil Biol 44:316–323

Morey KJ, Ortega JL, Gopalan SC (2002) Cytosolic glutamine synthetase in soybean is encoded by a multi gene family, and the members are regulated in an organ–specific and developmental manner. Plant Physiol 128:182–193

Ortiz N, Armada E, Duque E, Roldan A, Azcona R (2015) Contribution of arbuscular mycorrhizal fungi and/or bacteria to enhancing plant drought tolerance under natural soil conditions: effectiveness of autochthonous or allochthonous strains. J Plant Physiol 174:87–96

Porcel R, Ruiz-Lozano JM (2004) Arbuscular mycorrhizal influence on leaf water potential, solute accumulation, and oxidative stress in soybean plants subjected to drought stress. J Exp Bot 55:1743–1750. https://doi.org/10.1093/jxb/erh188

Porcel R, Barea JM, Ruiz-Lozano JM (2003) Antioxidant activities in mycorrhizal soybean plants under drought stress and their possible relationship to the process of nodule senescence. New Phytol 157:135–143. https://doi.org/10.1046/j.1469-8137.2003.00658.x

Purcell L, Serraj R, Sinclair T, De A (2004) Soybean N2 fixation estimates, ureide concentration, and yield responses to drought. Crop Sci 44:484–492

Rahmani H, Saleh-Rastin N, Khavazi K, Asgharzadeh A, Fewer D, Kiani S, Lindstrom K (2009) Selection of thermo tolerant Bradyrhizobial strains for nodulation of soybean (*Glycine max* L.) in semi–arid regions of Iran. World J Microbiol Biotechnol 25:591–600

Ramos MLG, Parsons R, Sprent JI (2005) Differences in ureide and amino acid content of water stressed soybean inoculated with *Bradyrhizobium japonicum* and *B. elkanii*. Pesq Agrop Brasileira 40:453–458

Reibach P, Streeter J (1983) Metabolism of 14C–labeled photosynthate and distribution of enzyme of glucose metabolism in soybean nodules. Plant Physiol 72:634–640

Rillig MC, Steinberg PD (2002) Glomalin production by an arbuscular mycorrhizal fungus: a mechanism of habitat modification. Soil Biol Biochem 34:371–1374. https://doi.org/10.1016/S0038-0717(02)00060-3

Rizhsky L, Liang H, Mittler R (2002) The combined effect of drought stress and heat shock on gene expression in tobacco. Plant Physiol 130:1143–1151

Ruiz-Lozano JM (2003) Arbuscular mycorrhizal symbiosis and alleviation of osmotic stress. New perspectives for molecular studies. Mycorrhiza 13:309–317

Ruiz-Lozano JM, Azcón R (1995) Hyphal contribution to water uptake in mycorrhizal plants as affected by the fungal species and water status. Physiol Plant 95:472–478

Ruiz-Lozano JM, Azcón R, Palma JM (1996) Superoxide dismutase activity in arbuscular mycorrhizal Lactuca sativa plants subjected to drought stress. New Phytol 134:327–333

Ruiz-Lozano JM, Collados C, Barea JM, Azcn R (2001) Arbuscular mycorrhizal symbiosis can alleviate drought induced nodule senescence in soybean plants. New Phytol 151:493–502

Saeki Y, Aimi N, Tsukamoto S, Yamakawa T, Nagatomo Y, Akao S (2006) Diversity and geographical distribution of indigenous soybean–nodulating bradyrhizobia in Japan. Soil Sci Plant Nutr 52:418–426

Safir GR, Boyer JS, Gerdemann JW (1972) Nutrient status and mycorrhizal enhancement of water transport in soybean. Plant Physiol 49:700–703

Sahgal M, Johri BN (2003) The changing face of rhizobial systematics. Curr Sci 84:43–48

Sanchez FJ, Manzanares M, de Andres EF, Tenorio JL, Ayerbe L (1998) Turgor maintenance, osmotic adjustment and soluble sugar and proline accumulation in 49 pea cultivars in response to water stress. Field Crop Res 59:225–235

Sandhya V, Ali SZ, Grover M, Reddy G, Venkateswarlu B (2009) Alleviation of drought stress effects in sunflower seedlings by exopolysaccharides producing *Pseudomonas putida* strain P45. Biol Fertil Soils 46:17–26

Serraj R, Sinclair TR, Purcell L (1999) Symbiotic N2 fixation response to drought. J Exp Bot 50:143–155

Sharma MP, Srivastava K, Sharma SK (2010) Biochemical characterization and metabolic diversity of soybean rhizobia isolated from Malwa region of central India. Plant Soil Environ 56:375–383

Sharp RE, Silk WK, Hsiao TC (1988) Growth of the maize primary root at low water potentials. I. Spatial distribution of expansive growth. Plant Physiol 87:50–57

Sharp RE, Poroyko V, Hejlek LG, Spollen WG, Springer GK, Bohnert HJ, Nguyen HT (2004) Root growth maintenance during water deficits: physiology to functional genomics. J Exp Bot 55:2343–2351

Smith SE, Read DJ (2008) Mineral nutrition, toxic element accumulation and water relations of arbuscular mycorrhizal plants. In: Smith SE, Read DJ (eds) Mycorrhizal symbiosis, 3rd edn. Academic, London, pp 145–118

Spaink HR (1994) The molecular basis of the host specificity of the Rhizobium bacteria. Antonie Van Leeuwenhoek 65:81–98

Sprent JI (1971) The effect of water stress on nitrogen–fixing root nodules. I. Effects on the physiology of detached soybean nodules. New Phytol 70:9–16

Streeter JG (2003) Effects of drought on nitrogen fixation in soybean root nodules. Plant Cell Environ 26:1199–1204

Streeter JG, Gomez ML (2006) Three enzymes for trehalose synthesis in *Bradyrhizobium* cultured bacteria and in bacteroids from soybean nodules. Appl Environ Microbiol 72(6):4250–4255

Streeter JG, Salminen SO, Whitmoyer RE, Carlson RW (1992) Formation of novel polysaccharides by Bradyrhizobium japonicum bacteroids in soybean nodules. Appl Environ Microbiol 58:607–613

Tilak KVBR, Saxena AK, Sadasivam KV (1995) Synergistic effects of phosphate–solubilizing bacterium (Pseudomonas striata) and arbuscular mycorrhizae on soybean. In: Adholeya A, Singh S (eds) Mycorrhizae: biofertilizer for the future. Tata Energy Research Institute, New Delhi, pp 224–226

Tisdall JM, Oades JM (1982) Organic matter and water–stable aggregates in soils. J Soil Sci 33:141–163. https://doi.org/10.1111/j.1365-2389.1982.tb01755.x

Uma C, Sivagurunathan P, Sangeetha D (2013) Performance of Bradyrhizobial isolates under drought conditions. Int J Curr Microbiol App Sci 2(5):228–232

Vaishnav A, Varma A, Tuteja N, Choudhary DK (2017) PGPR–mediated amelioration of crops under salt stress. In: Choudhary DK et al (eds) Plant–microbe interaction: an approach to sustainable agriculture. Springer Nature, Singapore, pp 205–226

Valentine AJ, Benedito VA, Kang Y (2011) Legume nitrogen fixation and soil abiotic stress: from physiology to genomics and beyond. Ann Plant Rev 42:207–248

Verbruggen E, Kiers ET (2010) Evolutionary ecology of mycorrhizal functional diversity in agricultural systems. Evol Appl 3:547–560

Vurukonda SSKP, Vardharajula S, Shrivastava M, Ali SKZ (2016) Enhancement of drought stress tolerance in crops by plant growth promoting rhizobacteria. Microbiol Res 184:13–24

Wang X, Pan Q, Chen F, Yan X, Liao H (2011) Effects of co-inoculation with arbuscular mycorrhizal fungi and rhizobia on soybean growth as related to root architecture and availability of N and P. Mycorrhiza 21:173–181. https://doi.org/10.1007/s00572-010-0319-1

Wilkinson S, Davies WJ (2010) Drought, ozone, ABA and ethylene: new insights from cell to plant to community. Plant Cell Environ 33(4):510–525

Willems A (2006) The taxonomy of rhizobia: an overview. Plant Soil 287:3–14

Wood JM, Bremer E, Csonka LN, Kraemer R, Poolman B, Heide TV, Smith LT (2001) Osmosensing and osmoregulatory compatible solute accumulation by bacteria. Comp Biochem Physiol Part A 130:437–460

Wright SF (2000) A fluorescent antibody assay for hyphae and glomalin from arbuscular mycorrhizal fungi. Plant Soil 226:171–177. https://doi.org/10.1023/A:1026428300172

Xie ZP, Staehelin C, Vierheilig H, Wiemken A, Jabbouri S, Broughton WJ, Vogeli-Lange R, Boller T (1995) Rhizobial nodulation factors stimulate mycorrhizal colonization of nodulating and nonnodulating soybeans. Plant Physiol 108:1519–1525

Xu ML, Ge C, Cui Z, Li J, Fan H (1995) Bradyrhizobium liaoningense sp. nov., isolated from the root nodules of soybeans. Int J Syst Bacteriol 45:706–711

Yang J, Kloepper JW, Ryu CM (2009) Rhizosphere bacteria help plants tolerate abiotic stress. Trends Plant Sci 14(1):1–4

Young JPW (1996) Phylogeny and taxonomy of rhizobia. Plant Soil 186:45–52

Zhang J, Jia W, Yang J, Ismail AM (2006) Role of ABA in integrating plant responses to drought and salt stresses. Field Crop Res 97(1):111–119

Chapter 8
In Silico Approach to Analyze the Biochemical Pathways of Bacterial Metabolite Synthesis

Tania, Mehendi Goyal, and Manoj Baranwal

Abstract Plant growth-promoting bacteria are well known to produce various bacterial secondary metabolites (BSM) which are much diversified in their origin and structure. These metabolites provide beneficial effect to crops which lead to sustainable agriculture. Each BSM is produced inside the bacterial cell by the regulation of specific biochemical pathways. To increase the yield of crops, there is a need to understand the metabolic pathway for the synthesis of secondary metabolites. Experimental approach to study the metabolic pathway was found to be limiting in terms of time and money. Development in the genome sequencing technologies and computational methods shifts the focus of researchers toward in silico approaches to reconstruct the biochemical pathways of metabolite production in genome-scale metabolic level. There are varieties of computational databases and tools available for the analysis of biosynthetic gene clusters including the prediction of non-ribosomal peptide synthetases (NRPS) and polyketide synthase (PKS) enzymes which encode for bacterial metabolites. These tools are developed to analyze the biochemical pathways for metabolite production. The present chapter gives a comprehensive overview on the in silico approach for the reconstruction of biochemical pathway and different databases and computational tools associated with it.

Keywords Biochemical pathways · Bacterial secondary metabolites · Biosynthetic gene clusters · Non-ribosomal peptide synthetases · Polyketide synthases · Genome mining

8.1 Introduction

Bacteria have been well known to enhance the crop plant growth and also safeguard plants from different biotic and abiotic factors. These bacteria are represented as plant growth-promoting bacteria (PGPB) and being commercially used as biocontrol

Tania · M. Goyal · M. Baranwal (✉)
Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala, India

D. K. Choudhary et al. (eds.), *In Silico Approach for Sustainable Agriculture*,
https://doi.org/10.1007/978-981-13-0347-0_8

agents, biofertilizers, bioprotectants, biosurfactants, biostimulants, and biotransformants (Vessey 2003; Compant et al. 2005; Odoh 2017). *Acetobacter*, *Azospirillum*, *Azotobacter*, *Bacillus*, *Burkholderia*, *Clostridium*, *Klebsiella*, *Pseudomonas*, *Rhizobium*, *Streptomyces*, and *Xenorhabdus* are some of the well-known PGPB (Bashan and De-Bashan 2005). Metabolites produced by these bacteria have different beneficial effect on crops that contribute to sustainable agriculture (Kudoyarova et al. 2015). Auxins, abscisic acid, cytokinins, ethylene, gibberellins, indole-3-acetic acid, and spermidine are different plant growth regulators produced by plants which help in elongation and cell division of roots and shoots (Gaspar et al. 1996). In the recent past, it has been studied from pragmatic literature survey that PGPB also produces these growth regulators as secondary metabolites, which are bacterial products, through specific biochemical pathways (Barea et al. 1976).

Bacteria's ability to grow and survive in harsh conditions is associated with the production of metabolites (Seaver et al. 2015). Many bacteria interact with plants through different types of association and produces secondary metabolites in their idiophase (stationary phase) (Souza et al. 2015). The secondary metabolites are not directly useful for bacteria but found to be associated with plant growth and development. These bacterial secondary metabolites (BSM) are commercially used to enhance the yield of crop such as improvement of nutrient acquisition and increase in the nutrient uptake capacity (phosphorous, organic compounds, and minerals), induce resistance, and provide antibiosis, growth promotion, and soil acidification to release organic acids and protons (Fig. 8.1) (Schippers et al. 1987; Matiru and Dakora 2004).

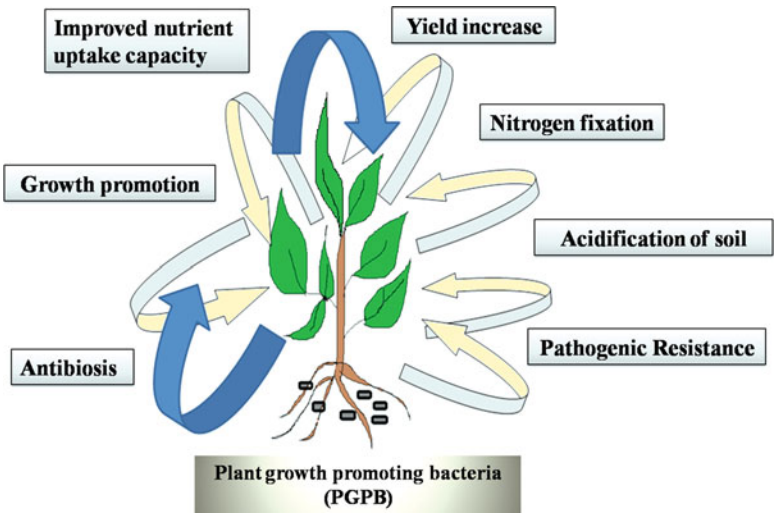


Fig. 8.1 Benefits of plant growth-promoting bacteria (PGPB) to plants

8.2 Bacterial Secondary Metabolites in Sustainable Agrosystems

Bacterial secondary metabolite (BSM) production in the bacteria operates under different anatomical, environmental, genomic, morphogenetic, stress, or temporal conditions (Wang et al. 2012). These BSMs can be produced by bacteria as endophytic (Brader et al. 2014; Jasim et al. 2014), entomopathogenic (Bode 2009), or halogenic (van Pée 1996). Several studies have been carried out on BSM production which shows that BSMs are produced inside a bacterial cell by the regulation of specific biochemical pathway and play a certain role in plant growth and development (Table 8.1).

Indole-3-acetic acid (IAA), cytokines, and gibberellins are different metabolites produced by *Rhizobium*, a potent nitrogen-fixing plant growth-promoting rhizobacteria (PGPR), which is being used as an eco-friendly and better alternative for chemical fertilizers in the agriculture and helps plants in development and killing of pathogens and weeds without harming the ecosystem (Gupta et al. 2015). Overproduction of indole-3-acetic acid, cytokinin, and mainly ethylene is found to be associated with the formation of crown gall in plants. 1-Aminocyclopropane-1-carboxylate (ACC) deaminase enzyme, a bacterial metabolite, is found to lower the ethylene level in plants by degrading ACC, a precursor in the ethylene production pathway (Fig. 8.2). In a study, ACC deaminase is isolated from *Pseudomonas putida* UW4 and introduced into *Agrobacterium tumefaciens* C58 for transfection and observed the inhibition of tumor growth in tomato and castor bean plants (Hao et al. 2007). In addition, ammonia produced as a by-product of ACC deaminase cycle is helpful for elongation of plant cells and nitrogen fixation.

A different category of BSMs are siderophores, which are fluorescent pigments produced by nitrogen-fixing bacteria such as *Rhizobium*, *Streptomyces*, and *Thermobifida*, that benefit the plant by chelating the iron groups (Franco-Correa and Chavarro-Anzola 2016). Siderophores are classified into three groups which are known as catecholates, hydroxamates, and hydroxide-carboxylates. Siderophores chelate the iron group by reducing Fe^{+3} to Fe^{+2} by the activity of nitrogenase enzyme, to get them available for different biological processes (Fig. 8.2). Enzyme nitrogenase requires 36 iron atoms to function properly for these activities. Antibiotics like alchivemycin A, ascomycin, clavulanic acid, istamycin, spirotryprostatin A, and herbicides are other metabolites which are produced by *Streptomyces*, which help plants to fight against different disease-causing bacteria and fungus (Fig. 8.2) (Bibb 2005; Chaudhary et al. 2013; de Jesus Sousa and Olivares 2016).

There is a need to enhance the production of BSM for commercial utilization and improvement of the agrosystems to fulfill the requirement of crop food availability worldwide. Several strategies have been employed for the enhancement purpose which includes addition of organic additives (Chen et al. 2000), stress conditions (Radhakrishnan et al. 2017), heterologous expression method (Tatsis and O'Connor 2016), and genetic and metabolic engineering (Kumar and Prasad 2011). Genome

Table 8.1 Major biochemical pathways regulation in bacteria for the production of different metabolites that benefits plants in different ways

Major pathway	Metabolite	Role	Bacteria (plant association)	References
Indole-3-pyruvate (IPA)	Indole-3-acetic acid	Promotes cell division and cell enlargement, stem elongation, root growth inhibition, increases storage of solutes inside the cells, herbicides	*Azospirillum* (free-living)	Spaepen and Vanderleyden (2011)
			Bacillus, *Burkholderia*, *Enterobacter*, *Pseudomonas*, *Streptomyces*, (endophytic)	
			Rhizobium (symbiotic)	
Plastid deoxyxylulose 5-phosphate pathway (PdPP)	Gibberellins	Plant growth and stem elongation, root growth, elongation of the internodes	*Azospirillum* (free-living)	Bottini et al. (2004)
			Bacillus (endophytic)	
			Rhizobium (symbiotic)	
AcdS gene expression (*AcdS* GE)	ACC deaminase, ammonia	Lower plant ethylene levels, nitrogen fixation, stress response	*Bacillus*, *Burkholderia*, *Enterobacter*, *Pseudomonas* (endophytic)	Singh et al. (2015)
Isochorismate synthase (ICS)	Siderophore	Iron chelation, mineral solubilization	*Burkholderia*, *Enterobacter*, *Pseudomonas*, *Streptomyces*, *Serratia* (endophytic)	Tounekti et al. (2013)
			Rhizobium (symbiotic)	
Cycloheximide	Cycloheximide, antibiotics (neutramycin and hygromycin A)	Inhibit the loss of chlorophyll in plants, defense mechanism, inhibitory to protein synthesis	*Streptomyces* (endophytic)	Stulberg et al. (2016)

sequence of many bacterial species is known which has opened the gateway for the enhanced production of BSM by studying and manipulating (Schuster et al. 1999) the genetic system of the bacteria (Stephanopoulos and Vallino 1991).The methods include the experimental study of gene addition and deletion, network robustness (study of overall network function by increased or decreased activity of enzymes), and optimal growth patterns (Price et al. 2003).

The major perplexity occurred in isolation and enhanced production of these BSMs was to elucidate the biochemical pathways responsible for their synthesis inside the bacterial cells and their transportation outside the cell. The upgrowth of

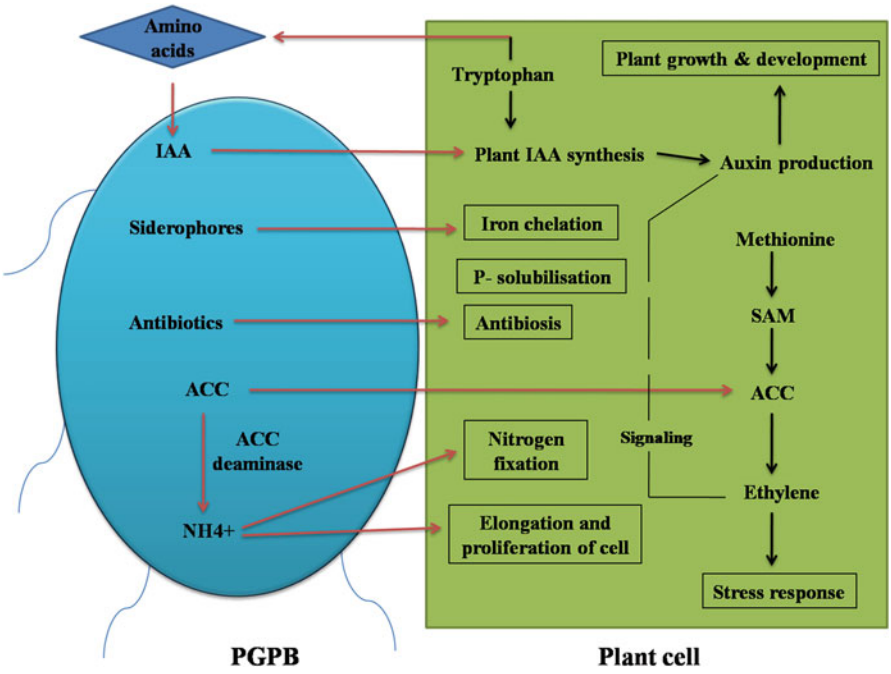


Fig. 8.2 Metabolite production in plant growth-promoting bacteria (PGPB) and their involvement and role in plant growth and development. *IAA* indole-3-acetic acid, *ACC* 1-aminocyclopropane-1-carboxylate, *SAM* S-adenosyl-L-methionine, and *P-solubilization* phosphate solubilization

system biology studies based on in silico approaches was the major breakthrough in understanding the incomplete genome-scale metabolic model. In silico approaches act as powerful tool to provide the contingency for large-scale collection of genome-scale metabolic models (Liu et al. 2014). These metabolic models with biochemical networks generate a pathway to study the molecular basis behind the production of different metabolites and their applications in agrosystems. The classical strategy adapted for the development of genome-scale metabolic models is to concatenate the generic models to the structured transcriptional regulatory network (TRN) models which include the understanding of cellular boundaries, enzyme-catalyzing biochemical reactions, exchange of fluxes with the external environment, and various signaling events in the biochemical pathway (Wang et al. 2012).

8.3 Biochemical Pathways in Metabolite Production

A metabolite is produced through a series of biochemical reactions that occur inside a living cell generating the network of biochemical pathways. This reaction occurs via metabolic flux generating step in which the product of one metabolic network

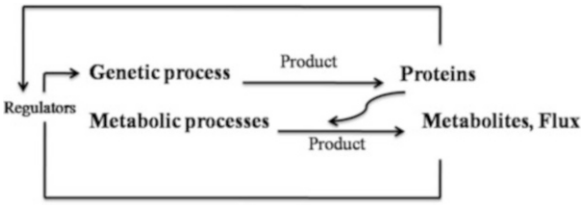


Fig. 8.3 Flow of genetic and metabolic processes inside the bacterial cell. The formed products (proteins and metabolites) and fluxes serve as the invader and regulators for the continuation of pathways

leads to the initiation of another metabolic network for the completion of pathway (Fig. 8.3). In each reaction, these metabolic steps result in the altered biochemical properties of the pathway and the formation of product (Edwards et al. 2002). Metabolic flux is the sum of all biochemical reactions that occur in the pathway through exchange of energy and mass. It is the measure of the rate of molecules turnover present in the metabolic pathway. Flux generation is interpreted by flux balance analysis (FBA) and metabolic flux analysis (MFA) method, in which flux gives the activity of metabolic network through the movement of mass within the metabolic networks connected by various cofactors (amino acids) and metabolites (Edwards and Palsson 2000).

Biochemical pathway study reveals the metabolic sequence and the presence of massive portion of putative biosynthetic gene clusters in the genome of organism which encodes for secondary metabolites. These pathways are operated under different committal, epigenetic, morphogenetic, or stress conditions to fulfill the requirement of desired metabolite in the cell. Indole-3-pyruvate (IPA) pathway is one of the examples of metabolic pathway (Fig. 8.4) which is regulated in many PGPBs (*Azospirillum*, *Bacillus*, *Enterobacter*, *Pseudomonas*, and *Rhizobium*) and leads to the synthesis of IAA (auxin) (Spaepen and Vanderleyden 2011). The key amino acid of the pathway is tryptophan which gets transaminated to IPA by the activity of enzyme aminotransferase. Then, this IPA is converted by indole-3-pyruvate decarboxylase (IPDC) to indole-3-acetaldehyde (IAA-ld) by decarboxylation reaction, which is finally oxidized to IAA by the activity of dehydrogenase. The IPDCs are encoded by the gene (ipdC) and are best characterized in *Enterobacter* (Spaepen and Vanderleyden 2011). Different BSM production in these PGPBs occurs simultaneously though the regulation of biochemical pathways and incorporates into the plant cell for various roles and functions (Tak et al. 2013; Glick 2014; Gamalero and Glick 2015).

8.3.1 Approach to Predict Biochemical Pathways

To know the metabolite production, it is of utmost importance to study the biochemical pathway which includes biochemical reactions and intermediate metabolite

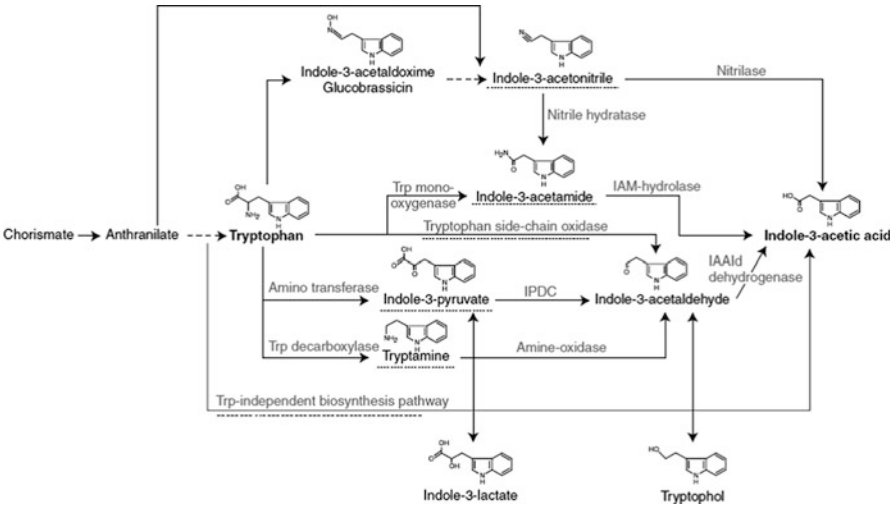


Fig. 8.4 Regulation of IPA (indole-3-pyruvate) pathway for the production of IAA (indole-3-acetic acid). (Adapted from Spaepen 2011)

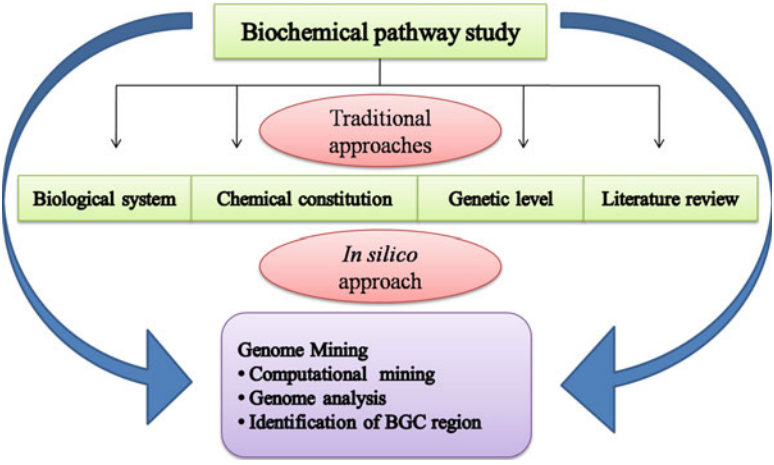


Fig. 8.5 Approaches employed in the development of biochemical pathways of metabolite production in bacteria

production. Traditionally, metabolites and their biochemical pathways were identified using different approaches (Weber 2014) (Fig. 8.5) which include (1) synthesizing metabolites in a biological system and studying enzyme-catalyzed biochemical reactions which lead to the generation of a defined pathway and (2) genetic approaches which include the use of conserved DNA probes for biosynthetic enzymes to identify the putative biosynthetic gene clusters (BGC) in genome sequence of the producing bacteria. The annotated genetic information is then

used for the cultivation of the producing bacteria (3) prediction of metabolites based on their chemical constitution and the knowledge of literature available. These approaches were found to be very laborious, costly, and time-consuming.

Advancement in the development of sequencing technologies and computational biology has made easier to generate the biochemical pathways of any metabolite. Genome mining approach was developed to predict metabolite production in the bacteria based on gathered genetic information without knowing the structure of the metabolite (Weber 2014). This method hastens the discovery of metabolite production by using various computational and bioinformatics tools.

8.3.1.1 In Silico Approach to Study Biochemical Pathway for BSM Synthesis

In silico approach targets the genetic and metabolic pathways to provide the knowledge about the genes and biosynthetic gene clusters (BGC) involved in the pathway. BGC regions contain those genes in the genome of bacteria which encode for biosynthesis enzymes, non-ribosomal peptide synthetase (NRPS), and polyketide synthase (PKS). These are the essential modular enzymes for the biosynthesis of secondary metabolites in bacteria (Kim et al. 2016). Several BGCs have been discovered for bacterial metabolites, but still the complete knowledge and function about their diversity are not known (Cimermancic et al. 2014).

Genome mining, the method developed to discern the genetic capability of living organism for the production of metabolites and novel compounds, is one of the important steps during the reconstruction of biochemical pathway. Genome mining involves the trial-and-error approaches in synthetic biology studies. This method firstly annotates the genome sequence of the bacteria using bioinformatics tool (BLAST and FASTA) to identify the putative BGC regions (Fig. 8.6). Then the heterologous expression of the identified BGC region is carried out to check the expression of the targeted genes and the production of the specific metabolite (Bachmann et al. 2014; Kim et al. 2016; Niu 2017) (Fig. 8.6). Once the information of synthesized product and the enzyme involved is known, it becomes easy to reconstruct the hypothetical biochemical pathway of the metabolite. To perform these target-based tasks, specialized bioinformatics softwares and tools are designed which generate and execute the biochemical pathway of metabolite production in the standard format (Weber 2014). Thus, this method predicts the metabolite and the biochemical reactions with fluxes involved in the pathway.

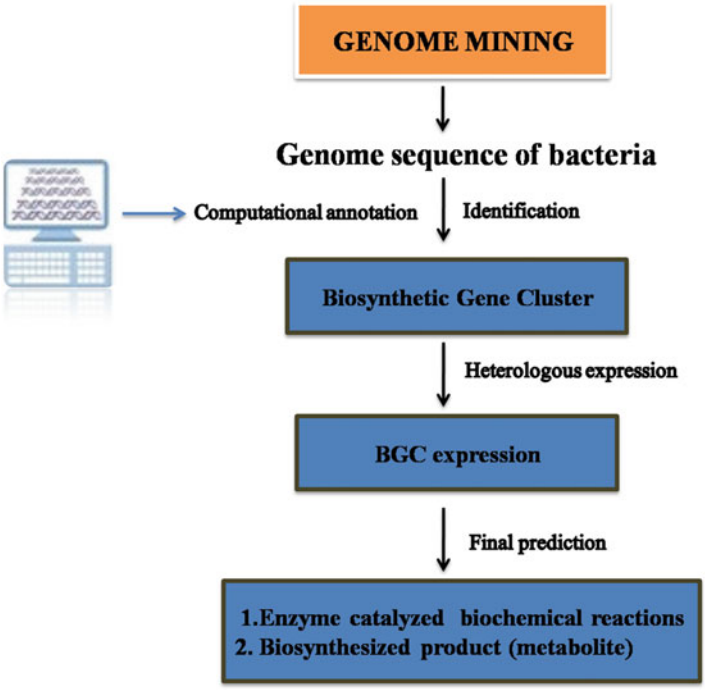


Fig. 8.6 Flowchart of genome mining method leads to the prediction of biochemical reactions and metabolite production

8.3.2 Reconstruction of Biochemical Pathway for Bacterial Metabolic Model

Reconstruction of the stoichiometry-based genome-scale metabolic model through in silico simulations has proven to be an effective approach in analyzing biochemical pathway of different bacterial metabolite synthesis. The analysis of genome annotation and metabolic biochemistry has enabled the programmer to reconstruct the bacterial metabolic model. These stoichiometric models provide the description about the regulation of biochemical pathways with the flow of energy and mass (Taffs et al. 2009). The study includes the computational simulation through the measure of mass flux balances of participating metabolites and stoichiometric coefficients.

To reconstruct the metabolic pathways, there is a need to study the cell physiology which includes the cellular boundaries, genome annotation, and metabolic biochemistry of the cell (Fig. 8.7) (Covert et al. 2001).

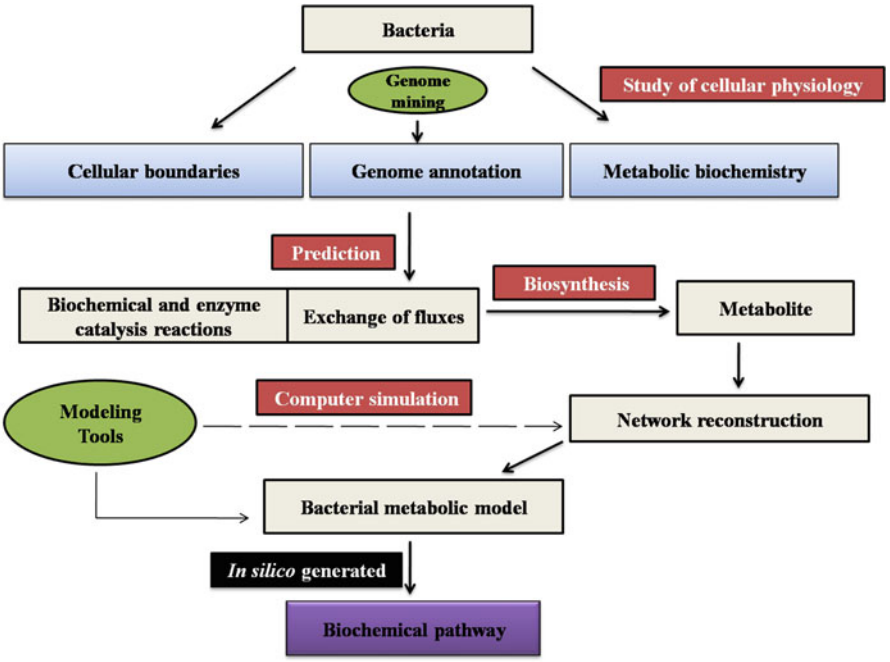


Fig. 8.7 Steps of biochemical pathway reconstruction using various in silico tools

8.3.2.1 Cellular Boundaries

It includes the understanding of bacterial cell physiology experimentally or theoretically that contributes to the identification of metabolic genes essential to regulate the metabolic pathway of bacteria. The physiological knowledge of bacterial cell is important to study because it predicts the presence or absence of certain biochemical reactions in the hypothetically generated biochemical pathway. For example, the experiment conducted on the growth pattern of bacteria has shown that a particular bacteria can survive without the production of a certain metabolite. Based on this experiment, the metabolic pathway of that bacteria is reconstructed which does not include the metabolic step of that metabolite production, without which the bacteria can survive. Although that metabolite is not crucial for survival of bacteria, it is produced in it which have some other benefits. The missing step of particular metabolite production should also be included in generated in silico hypothetical metabolic pathway for having real practical implementation, and a fully analyzed metabolic model can be generated. Once the physiological study is completed, the genome of the bacteria is analyzed to study essential genes that code for biochemical reaction catalyzing enzymes (Covert et al. 2001).

8.3.2.2 Genome Annotation

The first step in the computational annotation of genome is the identification of the coding regions or the open reading frames (ORF) in the genome sequence of bacteria. It includes the identification of the DNA sequence (BGC region) which encodes for PKS and NRPS enzymes with the help of genome mining method (Durot et al. 2008; Ziemert et al. 2016). The identification of each coding region is searched against the databases (BLAST and FASTA). They give the prediction of similarity between the query sequence and the deposited sequence in the database. As the number of sequences increased, the putative gene function and the structure of metabolite are determined by using different types of gene clustering databases (ClustScan, recombinant ClustScan, ClusterMine360, and DoBISCUIT).

8.3.2.3 Metabolic Biochemistry

After successful genome annotation, the study of metabolic biochemistry is done for three main purposes: (1) to assign particular function to the enzyme found in the genome, (2) to determine the presence of those missing biochemical reactions in the genome which have been excluded from the current genomic data, and (3) to validate the information found in the genome. There are two types of metabolic databases available to study metabolomics, one is the general purpose (Kyoto Encyclopedia of Genes and Genome (KEGG)) and metabolic pathway database (MPW) (Covert et al. 2001) and the other one is organism specific (EcoCyc) (Haggart et al. 2011).

8.3.2.4 Model Analyses

Once all the information of cell physiology is studied and validated, the gene clusters of PKS and NRPS enzymes, biochemical reactions with the exchange of fluxes, and enzymes involved are predicted. These information leads to the reconstruction of the metabolic network showing the biosynthesis of a particular metabolite. Computer simulation and various mathematical models (constraint-based model) are then used to validate the working and integrated functions of the reconstructed bacterial metabolic model. The behavior and analyses of generated pathway depend upon the factors like substrate availability, temperature, metabolic fluxes in the network, etc. Flux balance analysis (FBA) and metabolic flux analysis (MFA) are the flux analysis methods to describe the pathway working (Price et al. 2004) and are based on the principle of mass conservation for the metabolites in a given pathway, which states that "the overall production rate of a metabolite (R_p) is equal to the sum of the rates of the biochemical reactions producing it (R_r) and the associated stoichiometric coefficients (R_s) (Durot et al. 2008)."

$$R_{p=}R_r + Rs$$

8.3.2.5 Constraint-Based Mathematical Modeling

Constrain-based modeling (CBM) is the computational way of mathematical modeling to encode for the reconstruction of metabolic networks. CBM gives the insight of the complex system of metabolomics (Llaneras and Picó 2008) of bacteria by exploiting the properties of those biochemical reactions which have defined products and substrates (Conde et al. 2016). This method is useful for the prediction of biochemically feasible metabolic fluxes of steady state in bacterial cells (Little et al. 2005). CBM approach proceeds in a multistep procedure. In the first step, after the successful reconstruction of bacterial metabolic network, the regulatory association between the networks is developed. In the second step, the constraint-based statements are studied under which the network operates. The constraints are the different variables studied in the system as stoichiometric matrices (S) containing rows as unique metabolites (M) and columns as biochemical reactions with fluxes (N) and enzyme capacity. These statement-based constraint leads to the interpretation of a defined solution space which describes the functions of the reconstructed network. As the biochemical reactions yield the specific metabolite, S becomes the space matrix, and its stoichiometric coefficients can be tabulated as M × N matrix. In the third and last step, the practical significant and meaningful region in the space which corresponds to the possible solution is determined experimentally (Price et al. 2003).

Mass balance equation for CBM:

$$dx/dt = Sv \tag{8.1}$$

where x is the concentration vector of all metabolites with length m, and v is the flux distribution vector with length n in the biochemical reactions over the time t in the system. Now CBM is assumed to be operated under quasi-steady-state condition (in which the concentration of the metabolite does not change with time), and then $dx/dt = 0$; therefore, Eq. (8.1) becomes

$$Sv = 0$$

The most common approach to CBM for the identification of flux distribution vector (v) is the use of FBA method. FBA method establishes the relationship between the constraint-based analyses and enzymatic regulation in the biochemical reactions. It tells that the study of metabolic network with stoichiometric reactions is necessary to determine the degradation and synthesis of basic metabolites and specific genes whose products are allied with these enzymatic biochemical reactions. The first CBM model was developed which focuses on central carbon metabolism, but it was studied that the requirement of steady-state assumption of FBA did not comply with in silico modeling of bacteria in all the situations, especially during the

"diauxic shift" from one carbon source to the other carbon source. To overcome this problem, dynamic flux balance analysis (DFBA) was developed, which implements the programming of both dynamic (nonlinear) and static (linear) optimization of constraints with the rate of change of flux (Mahadevan et al. 2002).

8.3.3 In Silico Reconstructed Metabolic Model

The completed and most detailed genome-scale metabolic model of PGPB *Escherichia coli* (iJE660) (Souza et al. 2015) was reconstructed by Edwards and Palsson in 2000. The details of this model are stored in EcoCyc database with its reconstructed latest version of *i*JO1366 predicting 2251 biochemical reactions and 1136 metabolites which accounts for 1366 genes (Weaver et al. 2014). Genome-scale bacterial metabolic models have been also reconstructed for *Bacillus* sp., *Pseudomonas aeruginosa*, *Streptomyces lividans*, and *S. tsukubaensis* (Haggart et al. 2011; Kim et al. 2016).

8.4 In Silico Databases for the Prediction of BSM and Their Pathways

8.4.1 Databases on Biosynthetic Gene Clusters (BCG)

8.4.1.1 ClustScan and Recombinant ClustScan Database

The ClustScan database (CSDB) and recombinant ClustScan database (rCSDB) are in silico designed databases (Starcevic et al. 2008) for the rapid annotation of genome (DNA sequences) which encode for non-ribosomal peptide synthetases (NRPS), polyketide synthases (PKS), and the hybrid PKS/NRPS. CSDB database has currently a total of 170 gene clusters which include 51, 57, and 62 for NRPS, PKS, and hybrid PKS/NRPS, respectively. Actinomycetes, myxobacteria, and *Bacillus* species are being used to obtain these gene clusters (Diminic et al. 2013). The rCSDB has the information about the generated recombinants between PKS clusters, and it contains 20,187 recombinant gene clusters.

8.4.1.2 ClusterMine360 Database

The database of ClusterMine360 is a database organized on two key elements, the compound family (grouping of compounds having similar core structure) and gene clusters. The database contains >200 NRPS and PKS gene clusters and >185 compound families (Conway and Boddy 2012). This database retrieves the sequence

from the NCBI and analyzes it with antiSMASH (genome mining tool to predict BGC region), when the user uploads a particular NRPS or PKS cluster.

8.4.1.3 Biosynthesis Clusters Curated and Integrated (DoBISCUIT)

DoBISCUIT is a literature-based and manually curated database which focuses on NRPS and PKS gene clusters (Ichikawa et al. 2012). This database is developed using actinomycete bacteria. It has registered 72 gene clusters comprising 119 International Nucleotide Sequence Database Collection (INSDC) entries based on 516 collected references.

8.4.2 Tools for Analyzing PKS and NRPS Pathways

8.4.2.1 SEARCHPKS/NRPS-PKS/SBSPKS

SEARCHPKS is one of the tools available for characterization of useful domain in type 1 PKS acyltransferase domains (Yadav et al. 2003). When SEARCHPKS was integrated with the NRPS-PKS system, the diversity of its prediction also gets broadened for iterative type II and III PKS and NRPS. Recently, NRPS-PKS system is augmented by SBSPKS (a structure-based component) which gives the prediction of the docking interfaces and 3D homology modeling of PKS modules and domains. It identifies various PKS domains in a polypeptide sequence using automated computational protocol.

NP.searcher

NP.searcher is a HMM (hidden Markov model)-based web tool for the automatic annotation of genome to identify genes for PKS, NRPS, and hybrid PKS/NRPS enzyme system. It gives the prediction about the biosynthetic product and cyclic NRPS products. This tool fastens the screening of suitable natural product drug candidates utilizing available genomic information (Li et al. 2009).

8.4.2.2 MAPSI

MAPSI (Management and Analysis for Polyketide Synthase Type I) uses homology search and profile HMM (hidden Markov model) for identification of genes for PKS enzyme (Weber 2014). MAPSI integrates MapsiDB, a database which contains data on polyketides and its related genome. The web interface of MAPSI system consists of five parts which are MapsiDB, genome analysis, MAPSI tools, domain analysis, and management.

8.4.3 Tools for Genome Mining of Complete BGCs

8.4.3.1 BAGEL

BAGEL is a web-based mining tool developed for the identification of five main classes of peptides and bacteriocins (van Heel et al. 2013). These classes are (1) circular bacteriocins with carbohydrate moieties, (2) complex bacteriocins with lipid moieties, (3) lanthipeptides, which are post-translationally modified peptides (RIPPs), (4) heat stable non-modified bacteriocins, and (5) heat-labile large bacteriocins. All classes are identified using knowledge-based motif databases and bacteriocins databases. BAGEL contains RIPP database and identifies BGC region in genome to characterize its putative product.

8.4.3.2 PKMiner

PKMiner is a database tool that combines the capabilities of prediction of novel-type II PKS in the genome and aromatic polyketide chemotypes (Hwang et al. 2014). It works on the principle of aromatase and cyclase domain annotation. Currently, it provides the knowledge about more than 230 uncharacterized and 42 characterized type II PKS gene clusters of actinomycetes. The web interface of PKMiner is designed using Perl and Asynchronous Javascript and XML (AJAX).

8.4.3.3 CLUSEAN (Cluster Sequence Analyzer)

CLUSEAN is a Bioperl-based toolkit designed for genome annotation to identify BGC region of secondary metabolite synthesis in bacteria (Weber et al. 2009). It contains analyses section for automated homology search, classification of enzymes, and specificity predictions of conserved protein domains in NRPS, PKS, and NRPS A-domains. The annotation results of CLUSEAN are exported as EMBL OR MS Excel files.

antiSMASH (Antibiotics and Secondary Metabolite Analysis Shell)

antiSMASH is one of the widely used effective tools for genome mining of DNA sequences of bacteria for the identification of secondary metabolite BGCs region in genome (Weber et al. 2015). Its current version 3.0 is available for use, which was developed in a collaborative project between Tubingen University, UCSF, and Groningen University. Many tools and algorithms are operated by impairing with antiSMASH for the analyses of BGC region. This tool also contains a "ClusterBlast," and a "SubCluster Blast," for detecting conserved operons within BGC region. The annotated file can be exported as various file formats.

Table 8.2 An overview of the different databases, software, and tools available to predict biosynthetic gene clusters for the reconstruction of genomic-scale metabolic pathways

Program/database	URL
Databases on gene clusters	
ClusterMine360	http://www.clustermine360.ca/
ClustScan database	http://csdb.bioserv.pbf.hr/csdb/ClustScanWeb.html
DoBiSCUIT	http://www.bio.nite.go.jp/pks/
Recombinant ClustScan database	http://csdb.bioserv.pbf.hr/csdb/RCSDB.html
Software to analyze type I PKS and NRPS pathways	
ClustScan professional	http://bioserv.pbf.hr/cms/index.php?page=clustscan
MAPSI/ASMPKS	http://gate.smallsoft.co.kr:8008/pks/
NP.searcher	http://dna.sherman.lsi.umich.edu/
NRPS-PKS/SBSPKS/ SEARCHPKS	http://202.54.226.228/~pksdb/sbspks_updated/master.html
Tools for mining of secondary metabolite gene clusters	
antiSMASH 2	http://antismash.secondarymetabolites.org
BAGEL3	http://bagel2.molgenrug.nl/
CLUSEAN	https://bitbucket.org/antismash/clusean
MIDDAS-M	http://133.242.13.217/MIDDAS-M/
PKMiner	http://pks.kaist.ac.kr/pkminer/
SMURF	http://jcvi.org/smurf/index.php
ARTS	https://arts2.ziemertlab.com/
CASSIS and SMIPS	https://sbi.hki-jena.de/cassis/

These software/tools mainly focus on the characterization and identification of PKS and NRPS by decoding their characteristic domain architecture based on mainly HMM and homology search blueprint. In Table 8.2, different databases and tools for the prediction of BSM pathways are mentioned along with their URL.

8.5 Conclusion

The progress in sequencing technologies permitted the enhanced production of bacterial metabolites based on in silico approaches, which includes the major method of genome mining. The basic principle behind the genome mining approach is to identify the putative gene clusters in the genome of bacteria and predict the metabolite synthesis for the identified biochemical pathway. With the development of various computational tools and databases, it has become easier to analyze the biochemical pathway underlying bacterial metabolite synthesis.

References

Bachmann BO, Van Lanen SG, Baltz RH (2014) Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? J Ind Microbiol Biotechnol 41(2):175–184

Barea J, Navarro E, Montoya E (1976) Production of plant growth regulators by rhizosphere phosphate-solubilizing bacteria. J Appl Microbiol 40(2):129–134

Bashan Y, De-Bashan L (2005) Plant growth-promoting. Encycl Soils Environ 1:103–115

Bibb MJ (2005) Regulation of secondary metabolism in streptomycetes. Curr Opin Microbiol 8 (2):208–215

Bode HB (2009) Entomopathogenic bacteria as a source of secondary metabolites. Curr Opin Chem Biol 13(2):224–230

Bottini R, Cassán F, Piccoli P (2004) Gibberellin production by bacteria and its involvement in plant growth promotion and yield increase. Appl Microbiol Biotechnol 65(5):497–503

Brader G, Compant S, Mitter B, Trognitz F, Sessitsch A (2014) Metabolic potential of endophytic bacteria. Curr Opin Biotechnol 27:30–37

Chaudhary AK, Dhakal D, Sohng JK (2013) An insight into the "-omics" based engineering of streptomycetes for secondary metabolite overproduction. BioMed Res Int 2013

Chen G, Li X, Waters B, Davies J (2000) Enhanced production of microbial metabolites in the presence of dimethyl sulfoxide. J Antibiot 53(10):1145–1153

Cimermancic P, Medema MH, Claesen J, Kurita K, Brown LCW, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 158(2):412–421

Compant S, Duffy B, Nowak J, Clément C, Barka EA (2005) Use of plant growth-promoting bacteria for biocontrol of plant diseases: principles, mechanisms of action, and future prospects. Appl Environ Microbiol 71(9):4951–4959

Conde M, do Rosario P, Sauter T, Pfau T (2016) Constraint based modeling going multicellular. Front Mol Biosci 3:3

Conway KR, Boddy CN (2012) Clustermine360: a database of microbial pks/nrps biosynthesis. Nucleic Acids Res 41(D1):D402–D407

Covert MW, Schilling CH, Famili I, Edwards JS, Goryanin II, Selkov E, Palsson BO (2001) Metabolic modeling of microbial strains in silico. Trends Biochem Sci 26(3):179–186

Diminic J, Zucko J, Ruzic IT, Gacesa R, Hranueli D, Long PF, Cullum J, Starcevic A (2013) Databases of the thiotemplate modular systems (csdb) and their in silico recombinants (r-csdb). J Ind Microbiol Biotechnol 40(6):653–659

Durot M, Bourguignon P-Y, Schachter V (2008) Genome-scale models of bacterial metabolism: reconstruction and applications. FEMS Microbiol Rev 33(1):164–190

Edwards JS, Palsson BO (2000) Metabolic flux balance analysis and the in silico analysis of *Escherichia coli* k-12 gene deletions. BMC Bioinforma 1(1):1

Edwards JS, Covert M, Palsson B (2002) Metabolic modelling of microbes: the flux-balance approach. Environ Microbiol 4(3):133–140

Franco-Correa M, Chavarro-Anzola V (2016) Actinobacteria as plant growth-promoting rhizobacteria. In: Actinobacteria-basics and biotechnological applications. InTech

Gamalero E, Glick BR (2015) Bacterial modulation of plant ethylene levels. Plant Physiol 169 (1):13–22

Gaspar T, Kevers C, Penel C, Greppin H, Reid DM, Thorpe TA (1996) Plant hormones and plant growth regulators in plant tissue culture. In Vitro Cell Dev Biol-Plant 32(4):272–289

Glick BR (2014) Bacteria with acc deaminase can promote plant growth and help to feed the world. Microbiol Res 169(1):30–39

Gupta G, Parihar SS, Ahirwar NK, Snehi SK, Singh V (2015) Plant growth promoting rhizobacteria (pgpr): current and future prospects for development of sustainable agriculture. J Microb Biochem Technol 7(2):096–102

Haggart CR, Bartell JA, Saucerman JJ, Papin JA (2011) Whole-genome metabolic network reconstruction and constraint-based modeling⋆. In: Methods in enzymology. Elsevier, Amsterdam, pp 411–433

Hao Y, Charles TC, Glick BR (2007) Acc deaminase from plant growth-promoting bacteria affects crown gall development. Can J Microbiol 53(12):1291–1299

van Heel AJ, de Jong A, Montalban-Lopez M, Kok J, Kuipers OP (2013) Bagel3: automated identification of genes encoding bacteriocins and (non-) bactericidal posttranslationally modified peptides. Nucleic Acids Res 41(W1):W448–W453

Hwang K-S, Kim HU, Charusanti P, Palsson BØ, Lee SY (2014) Systems biology and biotechnology of streptomyces species for the production of secondary metabolites. Biotechnol Adv 32 (2):255–268

Ichikawa N, Sasagawa M, Yamamoto M, Komaki H, Yoshida Y, Yamazaki S, Fujita N (2012) Dobiscuit: a database of secondary metabolite biosynthetic gene clusters. Nucleic Acids Res 41 (D1):D408–D414

Jasim B, Joseph AA, John CJ, Mathew J, Radhakrishnan E (2014) Isolation and characterization of plant growth promoting endophytic bacteria from the rhizome of *Zingiber officinale*. 3. Biotech 4(2):197–204

de Jesus Sousa JA, Olivares FL (2016) Plant growth promotion by streptomycetes: ecophysiology, mechanisms and applications. Chem Biolo Technol Agric 3(1):24

Kim HU, Charusanti P, Lee SY, Weber T (2016) Metabolic engineering with systems biology tools to optimize production of prokaryotic secondary metabolites. Nat Prod Rep 33(8):933–941

Kudoyarova G, Arkhipova T, Melent'ev A (2015) Role of bacterial phytohormones in plant growth regulation and their development. In: Bacterial metabolites in sustainable agroecosystem. Springer, Cham, pp 69–86

Kumar RR, Prasad S (2011) Metabolic engineering of bacteria. Indian J Microbiol 51(3):403–409

Li MH, Ung PM, Zajkowski J, Garneau-Tsodikova S, Sherman DH (2009) Automated genome mining for natural products. BMC Bioinforma 10(1):185

Little PF, Subramaniam S, Jorde LB, Dunn MJ (2005) Encyclopedia of genetics, genomics, proteomics and bioinformatics. Wiley, Hoboken

Liu M, Bienfait B, Sacher O, Gasteiger J, Siezen RJ, Nauta A, Geurts JM (2014) Combining chemoinformatics with bioinformatics: in silico prediction of bacterial flavor-forming pathways by a chemical systems biology approach "reverse pathway engineering". PLoS One 9(1):e84769

Llaneras F, Picó J (2008) Stoichiometric modelling of cell metabolism. J Biosci Bioeng 105 (1):1–11

Mahadevan R, Edwards JS, Doyle FJ III (2002) Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. Biophys J 83(3):1331–1340

Matiru VN, Dakora FD (2004) Potential use of rhizobial bacteria as promoters of plant growth for increased yield in landraces of African cereal crops. Afr J Biotechnol 3(1):1–7

Niu G (2017) Genomics-driven natural product discovery in actinomycetes. Trends Biotechnol 36 (3):238–241

Odoh CK (2017) Plant growth promoting rhizobacteria (pgpr): a bioprotectant bioinoculant for sustainable agrobiology. Rev Int J Adv Res Biol Sci 4(5):123–142

van Pée K-H (1996) Biosynthesis of halogenated metabolites by bacteria. Ann Revin Microbiol 50 (1):375–399

Price ND, Papin JA, Schilling CH, Palsson BO (2003) Genome-scale microbial in silico models: the constraints-based approach. Trends Biotechnol 21(4):162–169

Price ND, Reed JL, Palsson BØ (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Microbiol 2(11):886

Radhakrishnan R, Hashem A, Abd_Allah EF (2017) Bacillus: a biological tool for crop improvement through bio-molecular changes in adverse environments. Front Physiol 8:667

Schippers B, Bakker AW, Bakker PA (1987) Interactions of deleterious and beneficial rhizosphere microorganisms and the effect of cropping practices. Annu Rev Phytopathol 25(1):339–358

Schuster S, Dandekar T, Fell DA (1999) Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends Biotechnol 17(2):53–60

Seaver S, Bradbury LM, Frelin O, Zarecki R, Ruppin E, Hanson AD, Henry CS (2015) Improved evidence-based genome-scale metabolic models for maize leaf, embryo, and endosperm. Front Plant Sci 6:142

Singh RP, Shelke GM, Kumar A, Jha PN (2015) Corrigendum: biochemistry and genetics of acc deaminase: a weapon to "stress ethylene" produced in plants. Front Microbiol 6:937

Souza Rd, Ambrosini A, Passaglia LM (2015) Plant growth-promoting bacteria as inoculants in agricultural soils. Genet Mol Biol 38(4):401–419

Spaepen S, Vanderleyden J (2011) Auxin and plant-microbe interactions. Cold Spring Harb Perspect Biol 3(4):a001438

Stephanopoulos G, Vallino JJ (1991) Network rigidity and metabolic engineering in metabolite overproduction. Science 252(5013):1675–1681

Starcevic A, Zucko J, Simunkovic J, Long PF, Cullum J, Hranueli D (2008) Clustscan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. Nucleic Acids Res 36(21):6882–6892

Stulberg ER, Lozano GL, Morin JB, Park H, Baraban EG, Mlot C, Heffelfinger C, Phillips GM, Rush JS, Phillips AJ (2016) Genomic and secondary metabolite analyses of streptomyces sp. 2aw provide insight into the evolution of the cycloheximide pathway. Front Microbiol 7:573

Taffs R, Aston JE, Brileya K, Jay Z, Klatt CG, McGlynn S, Mallette N, Montross S, Gerlach R, Inskeep WP (2009) In silico approaches to study mass and energy flows in microbial consortia: a syntrophic case study. BMC Syst Biol 3(1):114

Tak HI, Ahmad F, Babalola OO (2013) Advances in the application of plant growth-promoting rhizobacteria in phytoremediation of heavy metals. In: Reviews of environmental contamination and toxicology, vol 223. Springer, New York, pp 33–52

Tatsis EC, O'Connor SE (2016) New developments in engineering plant metabolic pathways. Curr Opin Biotechnol 42:126–132

Tounekti T, Hernández I, Munné-Bosch S (2013) Salicylic acid biosynthesis and role in modulating terpenoid and flavonoid metabolism in plant responses to abiotic stress. In: Salicylic acid. Springer, Dordrecht, pp 141–162

Vessey JK (2003) Plant growth promoting rhizobacteria as biofertilizers. Plant Soil 255(2):571–586

Wang Y, Eddy JA, Price ND (2012) Reconstruction of genome-scale metabolic models for 126 human tissues using mcadre. BMC Syst Biol 6(1):153

Weaver DS, Keseler IM, Mackie A, Paulsen IT, Karp PD (2014) A genome-scale metabolic flux model of *Escherichia coli* k–12 derived from the ecocyc database. BMC Syst Biol 8(1):79

Weber T (2014) In silico tools for the analysis of antibiotic biosynthetic pathways. Int J Med Microbiol 304(3–4):230–235

Weber T, Rausch C, Lopez P, Hoof I, Gaykova V, Huson D, Wohlleben W (2009) Clusean: a computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. J Biotechnol 140(1–2):13–17

Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W (2015) Antismash 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res 43(W1):W237–W243

Yadav G, Gokhale RS, Mohanty D (2003) Searchpks: a program for detection and analysis of polyketide synthase domains. Nucleic Acids Res 31(13):3654–3658

Ziemert N, Alanjary M, Weber T (2016) The evolution of genome mining in microbes–a review. Nat Prod Rep 33(8):988–1005

Chapter 9
In Silico Modulation Techniques for Upgrading Sustainability and Competitiveness in Agri-food Sector

Nitika Thakur

Abstract Agricultural sustainability and competitiveness of organic markets are the interlinked areas which jointly needs a serious makeover. In silico approaches generally highlight an insight to the use of signature sequences (promoter elements) and other related flanking elements. Agricultural sector mainly organic sector needs to be linked with modern techniques which can elevate its level and provides a clear explanation about the drawbacks and sources of improvement. The following chapter deals with the association of modern techniques like nanotechnology which can be linked further with agricultural problems to elevate conditions like sustainability and competitiveness of organic markets. Taking into account the various deleterious effects of conventional farming, the need of the hour demands for a technological interventional related with eco-friendly techniques like "green technology," so that we can improve the status of organic markets and can easily understand the loopholes associated with traditional methodology which are followed by the farmers. Agriculture, natural resources, and food have been always linked to burning challenges (sustainability, susceptibility, human health). The aim of integrating techniques like using nanomaterials in agriculture is to bring reductions in levels of chemicals used, minimize nutrient losses during fertilization, and increase yield.

9.1 Introduction

Rigorous agricultural technologies are needed to equalize the food requirements for the uncontrollable world population.

However, it is strongly associated with the mass consumption of non-replenishable natural resources and the everlasting climate changes. The challenge in this era is to attain sustainable environmental issues with good yields. In this context, technologies like agro-nanotechnologies need to be enhanced.

N. Thakur (✉)
Department of Biotechnology, Shoolini University, Solan, India

D. K. Choudhary et al. (eds.), *In Silico Approach for Sustainable Agriculture*,
https://doi.org/10.1007/978-981-13-0347-0_9

In silico modulation techniques directly relate to the nano-world which finally focuses on:

1. Understanding mechanisms of root-microbiome interactions and highlighting the ways of strengthening these types of interactions
2. Possible ways to improve the soil microbe ability for studying the negative and positive impacts of stress factors on crop productivity
3. Encourage beneficial organisms while preventing the presence of pathogens
4. Ensuring biosafety of agricultural food products, so that they are safe for health

9.2 Green Synthesis

Green synthesis includes synthesis through plants, bacteria, fungi, algae, etc. In recent years, green synthesis of ZnO nanoparticles was achieved by using *Aloe vera* leaf extract (Sangeetha et al. 2011). *Aloe vera* has exceptional therapeutic properties. *Aloe vera* has been shown to have anti-inflammatory activity, immunostimulatory activity, and cell growth stimulatory activity. Activity against a variety of infectious agents in terms of antiviral, antifungal, and antibacterial has also been reported. Green synthesis procedures involve the plant-based synthesis of nanoparticles. Green synthesis techniques make use of somewhat pollutant-free chemicals for synthesis of nanostructures. It embraces the use of eco-friendly and safe solvents such as water and natural extracts.

So biological approaches using microorganisms and plants or plant extracts for synthesis of metal nanoparticles have been suggested as safe alternatives to chemical methods. In biogenic synthesis of nanoparticles, several biological systems including bacteria, fungi, and yeast have been used safely. But synthesis of nanoparticles by using microorganisms is somewhat difficult because it involves elaborate process of maintaining cell cultures, intracellular synthesis, and multiple purification steps.

Plant parts like roots, leaves, stems, seeds, and fruits have also been utilized for the NPs synthesis as their extract is rich in phytochemicals which act as both reducing and stabilization agent (Subramanian et al. 2016).

Much attention has been diverted to the green synthesis of metallic NPs using biological materials as reducing and stabilizing agents and due to the usage of eco-friendly, nontoxic, and safe reagents during the biosynthesis process; green synthesis has been considered safe in the field of toxic chemical and physical methods (Kasana et al. 2016).

In this biological method, plant extracts are used for the controlled and tedious synthesis of various metal NPs. A large fraction of surface atoms is responsible for atomic behavior of nanoparticles (Sidra et al. 2014). Despite the fact that conventional methods use less time to synthesize nanoparticles, contributing to environmental toxicity because they require toxic chemicals as leveling agents. Green nanotechnology is a respectful environmental friendly selection and is profitable and uses natural proteins as covered agents.

The literature states that some of the toxic chemicals that are used in physical and chemical methods may reside in the NPs formed which may prove to be hazardous in the field of their application (Chandran et al. 2006). Thus, an environment-friendly and cost-effective method is required for nanoparticle synthesis. Physical method involves the use of high vacuum in processes like pulsed laser deposition, MBE (molecular beam epitaxy), thermal evaporation, etc. (Yuvakkumar et al. 2015), and chemical process includes chemical microemulsion, wet chemical, spray pyrolysis, electrodeposition, chemical and direct precipitation, and microwave-assisted combustion (Yuvakkumar et al. 2015). Additional capping and stabilizing agent are needed in physical and chemical methods.

The nanoparticles synthesized by green synthesis are generally helpful in detecting problems in agricultural systems and enhancing better food quality so that the agricultural food becomes competitive in agri-food sectors. Therefore, this technology should be adopted that focuses on getting higher and sustainable agricultural production. The agriculture development also depends on the social inclusion, health, climate changes, energy, ecosystem channels, natural resources, etc., Therefore, sustainable agriculture requires stable environmental performance; participation of food chain ecosystems is, hence, required in relation to sustainable food production and enhancing competitive modes in food markets (Thornhill et al. 2016).

9.3 Role of Actinomycete-Mediated Nanosystem in Agriculture

Actinobacteria are specified group of microorganisms which share the common characteristic of both bacteria and fungi and are known to play a ubiquitous role in agricultural production systems. The important functions include the production of wide range of growth-promoting compounds and metabolites that assist the host plants to resist both biotic and abiotic stress conditions. Actinobacteria are often used as a biocontrol agent (BCA) against most of the plant pathogens.

On the other hand, actinobacteria colonize host plants and secrete growth-promoting substances that favor good growth of plants even under adverse environmental conditions (nutrient deficiencies, drought, saline conditions, and heavy metal-contaminated soils). Several actinobacteria help in mycorrhizal symbiosis and biological nitrogen fixation as these are involved in the nutrient solubilization and mobilization of phosphates and iron. A volatile compound called "geosmin" which acts as biological indicator is also secreted by these groups of these organisms. Keeping in view the above advantages recently, research reports have suggested that actinobacteria are capable of producing metal oxide nanoparticles that can be exploited in the green synthesis of nanomaterials and utilized in biological as well as agricultural systems (Hamdali et al. 2008).

9.4 Copper Nanoparticles in Agriculture: Biological Synthesis and Antimicrobial Activity

Copper forms an essential micronutrient which plays a significant role in the health as well as nutrition of plants. Copper nanoparticles due to unique properties are more efficient than bulk copper particles in activity and functioning. Due to antimicrobial activity, copper nanoparticles are finding new applications in agriculture, healthcare, and industry. However there are growing concerns regarding the indiscriminate use of either copper or copper nanoparticles which can cause toxic effects to plants and other living organisms (Atha et al. 2012).

Biologically synthesized copper nanoparticles show good antibacterial and antifungal activity inhibiting the growth of pathogenic bacteria belonging to Gram-positive and Gram-negative genera and plant pathogenic fungi. Growth inhibition has been seen in case *Staphylococcus aureus*, *Enterococcus faecalis*, *Propionibacterium acnes*, *Escherichia coli*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *Shigella flexneri*, *Proteus vulgaris*, and *Salmonella typhimurium*. Antifungal activity of copper nanoparticles against the pathogenic fungi *Fusarium culmorum*, *Fusarium oxysporum*, *Fusarium graminearum*, and *Phytophthora infestans* has also been recorded. Copper nanoparticles at concentrations below 100 ppm have been reported to enhance germination and growth of some plants.

9.5 Use of Nanosensors for Detecting Agricultural Food Products

Nanosensors have been proven successful in food labeling. Further, in combination with reactive nanolayers, it may be helpful in quick and efficient recognition of food product.

Nanobarcode which consists of different gold and silver stripes serves as printed labels in the packaging of food that indicate the following specifications highlighting temperature, time, pathogens, freshness, humidity, etc. Lastly, we can suggest that the contaminant or nutrient sorption on NPs surfaces has attracted the attention of researchers for more studies on soil chemistry, showing that NPs have high sorption capacities for metal and anionic contaminants (Li et al. 2007). A phenomenon with significant consequences for remediation processes can exhibit. Metallic species such as Ni can be linked to natural aluminosilicates, TiO2 surfaces, humic acids, and aromatic compounds, and these association can be considered as very potent source of bioremediation in nano-agricultural system.

9.6 Agri-food Packaging and Labeling

Quality, safety, taste, and freshness in the agri-food industry in whole supply chain are the important parameters in packaging and labeling products. Recently, some nanosensors have been incorporated to detect the oxidation process in packaging materials by indicating the color change in food package. This technology has been successfully applied in package of milk and meat and extends the shelf life of food products. In the food packaging industry, the most used materials are plastic polymers that can be incorporated or coated with nanomaterials for improved mechanical or functional properties (Berekaa 2015). Moreover, nanocoatings served on food contact surfaces act as barrier for disinfection of bins, thus minimizing harmful bacterial growth. The nanotechnology is a forward marching process, which directs toward agricultural biosecurity.

9.7 Applications of Nanotechnology in Agriculture

To enhance the agricultural sustainability and productivity, nanotechnology, by virtue of nanoparticles, has offered potential applications in agriculture sector that include nanofertilizer, nanopesticide, nanoherbicide, nanosensor, and SDS (smart delivery systems) for efficient release of agrochemicals (Salamanca-Buentello et al. 2005; Oliveira et al. 2014; Campos et al. 2014; Grillo et al. 2016).

The benefits of nanotechnology in agricultural sector have led to a great interest, as it can increase agricultural productivity with lesser input of cost and energy. Additionally, for plant breeding and genetic purposes, use of nanotechnology-based devices (Jiang et al. 2013) is encouraged (Ghormade et al. 2011; Kah and Hofmann 2014; Parisi et al. 2015; Mishra and Singh 2016). Nanotechnological and in silico approach is leading to extreme benefits in the agricultural sector in number of ways like encapsulation of pesticides in nanoparticles for their easy release, nanoparticle-based genetic delivery system for crop improvement, seed germination of rainfed crops by carbon-assisted nanotubes, nanofertilizer for enhancing the crop nutrition and productivity, use of alternative nanopesticide for plant disease management, and use of nanoherbicide for elimination of weeds and nanosensors for detection of pathogens and monitoring of soil (Li et al. 2007; Barik et al. 2008; Wilson et al. 2008). Besides these, the major impact of nanomaterials on fate and accumulation of co-contaminants is the recent development (Servin and White 2016).

Nanotechnology provides the direction towards precision farming in today's world, where increasing demand of sustainability stresses to reduce the indiscriminate use and cost agricultural and natural resources (Chen and Yada 2011).

9.8 Future Perspectives and Conclusion

The future technology "nanotechnology" possesses very specific property in food supply chain that specifies the areas like crop production, use of agrochemicals such as nanofertilizers and nanoherbicides, accurate farming techniques, intelligent feed, nanopesticides, increasing food texture and quality and nutrient values, packaging and labeling, etc. around the world agricultural sector.

The focused areas need more attention in the near future in the field of agri-nanotechnology or nanofoods. Thus, it is important to take a modernized knowledge in agriculture. In agriculture process, still developing countries are suffering from lack of knowledge highlighting importance of food products.

So, development of database and alarm system, in silico modulation techniques, and international cooperation for regulation and legislation are necessary for improving this technology.

References

Atha DH, Wang H, Petersen EJ, Cleveland D, Holbrook RD, Jaruga P, Dizdaroglu M, Xing B, Nelson BC (2012) Copper oxide nanoparticle mediated DNA damage in terrestrial plant models. Environ Sci Technol 46(3):1819–1827

Barik TK, Sahu B, Swain V (2008) Nanosilica-from medicine to pest. Parasitol Res 103:253

Berekaa MM (2015) Nanotechnology in food industry; advances in food processing, packaging and food safety. Int J Curr Microbiol App Sci 4:345–357

Campos EV, Oliveira JL, Fraceto LF (2014) Applications of controlled release systems for fungicides, herbicides, acaricides, nutrients, and plant growth hormones: a review. Adv Sci Eng Med 6:373–387

Chandran SP, Chaudhary M, Pasricha R, Ahmad A, Sastry M (2006) Synthesis of gold nanotriangles and silver nanoparticles using Aloevera plant extract. Biotechnol Prog 22 (2):577–583

Chen H, Yada R (2011) Nanotechnologies in agriculture: new tools for sustainable development. Trends Food Sci Technol 22:585–594. control. *Parasitol. Res* 103: 253–258

Ghormade V, Deshpande MV, Paknikar KM (2011) Perspectives for nanobiotechnology enabled protection and nutrition of plants. Biotechnol Adv 29:792–803

Grillo R, Abhilash PC, Fraceto LF (2016) Nanotechnology applied to bio-encapsulation of pesticides. J Nanosci Nanotechnol 16:1231–1234

Hamdali H, Hafidi M, Virolle M, Ouhdouch Y (2008) Rock phosphate-solubilizing actinomycetes: screening for plant growth-promoting activities. World J Microbiol Biotechnol 24:2565–2575

Jiang S, Eltoukhy AA, Love KT, Langer R, Anderson DG (2013) Lipidoid-coated iron oxide nanoparticles for efficient DNA and siRNA delivery. Nano Lett 13:1059–1064

Kah M, Hofmann T (2014) Nanopesticide research: current trends and future priorities. Environ Int 63:224–235

Kasana RC, Panwar NR, Kaul RK, Kumar P (2016) Copper nanoparticles in agriculture: biological synthesis and antimicrobial activity. In: Ranjan S, Dasgupta N, Lichtfouse E (eds) Nanoscience in food and agriculture 3, Sustainable agriculture reviews 23. Springer, Cham

Li ZZ, Chen JF, Liu F, Liu AQ, Wang Q, Sun HY (2007) Study of UV shielding properties of novel porous hollow silica nanoparticle carriers for avermectin. Pest Manag Sci 63:241–246

Mishra S, Singh HB (2016) Preparation of biomediated metal nanoparticles. Indian Patent Filed 201611003248

Oliveira JL, Campos EV, Bakshi M, Abhilash PC, Fraceto LF (2014) Application of nanotechnology for the encapsulation of botanical insecticides for sustainable agriculture: prospects and promises. Biotechnol Adv 32:1550–1561. organophosphorus pesticides enzyme biosensor based on magnetic composite nanoparticles modified screen printed carbon electrode. *Sensors* 10, 625–638

Parisi C, Vigani M, Rodriguez-Cerezo E (2015) Agricultural nanotechnologies: what are the current possibilities? Nano Today 10:124–127. https://doi.org/10.1016/j.bios.2015.11.086

Salamanca-Buentello F, Persad DL, Court EB, Martin DK, Daar AS, Singer PA (2005) Nanotechnology and the developing world. PLoS Med 2:e97

Sangeetha G, Rajeshwari S, Venckatesh R (2011) Green synthesis of zinc oxide nanoparticles by aloe barbadensis miller leaf extract: structure and optical properties. Mater Res Bull 46 (12):2560–2566

Servin AD, White JC (2016) Nanotechnology in agriculture: next steps for understanding engineered nanoparticle exposure and risk. NanoImpact 1:9–12

Sidra S, Arshad M, Chaudhari SK (2014) Zinc oxide nanoparticles for revolutionizing agriculture: synthesis and applications. Sci World J 2014:1–8

Subramanian KS, Muniraj I, Uthandi S (2016) Role of actinomycete-mediated nanosystem in agriculture. In: Subramaniam G, Arumugam S, Rajendran V (eds) Plant growth promoting actinobacteria. Springer, Singapore

Thornhill S, Vargyas E, Fitzgerald T, Chisholm N (2016) Household food security and biofuel feedstock production in rural Mozambique and Tanzania. Food Sec 8:953–971

Wilson MA, Tran NH, Milev AS, Kannangara GSK, Volk H, Lu GHM (2008) Nanomaterials in soils. Geoderma 146:291–302

Yuvakkumar R, Suresh J, Saravanakumar B, Nathanael AJ, Hong SI, Rajendran V (2015) Rambutan peels promoted biomimetic synthesis of bioinspired zinc oxide nanochains for biomedical applications. Spectrochim Acta A Mol Biomol Spectrosc 137:250–258

Chapter 10
Bioinformatic Tools to Study the Soil Microorganisms: An In Silico Approach for Sustainable Agriculture

Pankaj Bhatt and Anupam Barh

Abstract The twenty-first century is the era of omics technologies which is mainly focused on generation and analysis of molecular data present within the organisms. In the last two decades, enormous data were generated by researchers in laboratories, due to the rapid developments of high-throughput next-generation sequencing (NGS) technologies. These data generated by these technologies can directly be applied to the agricultural developments. The agriculture system which is directly connected to soil can act as plant growth promoters in free-living state or either associated with the rhizospheric region. Whole-genome sequences of the microorganisms are available in the database which is useful for genome-wide identification of specific genes, proteins, ESTs, ORFs, etc. Identification through DNA barcoding in soil microorganism is also a new avenue where various bioinformatic tool assisted the process like MUSCLE, BRONX, ecoPrimers, etc. Microbial system biology is another way to explore the data from different metabolic pathways, genes, and proteins for the valid conclusion of the microbial activity. In totality, the in silico tools comprised of databases and softwares that can assist to reduce the "sequence-function gap" and help in the broad-spectrum study of soil microorganisms and their application toward sustainable agriculture.

10.1 Introduction

Soil microbial diversity is concerned with the study of all biological aspect of the microorganism (bacteria, archaea, fungi, viruses, protozoa, nematode, insect, earthworm, and parasites) that exists in the soil environment. Development of new approaches for characterizing microbial communities and imaging soil environments has benefited soil microbiology by providing new ways of detecting and locating

P. Bhatt (✉) · A. Barh
Department of Microbiology, Dolphin (P.G) Institute of Biomedical and Natural Sciences, Dehradun, Uttarakhand, India

ICAR-Directorate of Mushroom Research, Solan, Himanchal Pradesh, India

D. K. Choudhary et al. (eds.), *In Silico Approach for Sustainable Agriculture*,
https://doi.org/10.1007/978-981-13-0347-0_10

microorganisms. Soil microbiology achieved success from traditional isolation methods to recent high-throughput methods. A system approach will enable the structures of microbial communities to be characterized and will inform how microbial communities affect soil function. System approaches require accurate analysis of the spatiotemporal properties of the different microenvironments present in soil (O'Donnell et al. 2007). Different soil layers contain contrasting microbial cells. Globally marine surface sediment constitutes habitat for estimated 1.7×1028 bacteria and archaea (Whitman et al. 1998). Bulk sediment to sand grain addresses the microbial community directly to its microhabitat. On the basis of 16S rRNA, microbial communities were characterized by using fluorescence in situ hybridization. Sand grain in the range of 202–635 μm diameter contains 10^4–10^5 cells. In soil, each sand grain harbors highly diverse bacterial communities, shown by several thousand species-level operational taxonomic units (OTUs). Four to eight single grains are required to cover 50% OTU richness found on bulk sediment (Probandt et al. 2017). The development of molecular tools for biodiversity characterization based on DNA extraction from the soil matrix applied so far mostly to microorganisms or from organisms initially extracted from soils mainly fauna but also microorganisms through previous in vitro cultivation represents unprecedented opportunities (Ogram 2000). DNA barcoding is another interesting tool to study the soil microbial communities nowadays. Due to large-scale development of next-generation sequencing (NGS) methods, millions of DNA sequences in a relatively shorter period of time are generated. Microbarcodes and mini-barcodes are the comparatively shorter sequence of about 25 bp and 100–300 bp, respectively. These barcodes overcome the limitation of poor DNA quality. Mini-barcode can capitalize the barcoding in much older sample, i.e., 200 years old than standard barcode which only covers 10 years life. Mini-barcode has another advantage that it can resolve 3–5% lower species than the standard one. Thus emergence micro- and mini-barcodes can be utilized for the aged sample, preserved sample, processed samples, pharmaceuticals, and environmental samples. Since these barcodes are capable of identification of sample even from the small sample of DNA, they thus require the sophisticated and high level of sequencing techniques like next-generation sequencing (Kress and Erickson 2012). Application of NGS technologies has resulted in an increasing number of metabarcoding surveys on soil biodiversity conducted in a wide range of environment, i.e., agricultural field, grassland, forests, Arctic or Antarctic, and desert (Nielsen and Wall 2013; Orgiazzi et al. 2015).

10.2 Soil Microorganisms and Agriculture

Soil contains a rich pool of nutrients (carbon, nitrogen, phosphorus, and potassium) and microorganisms. For the growth and development of crop plants, these nutrients are benefitted for the plant health. Microorganisms have the ability to transfer these

nutrients from the soil to plant via their metabolic activities. Nitrogen, phosphorous, and carbon are the major component to be transferred through the microbial metabolisms. Bacteria associated with the rhizospheric soil of crop plants are helpful to provide nutrition through root system; such type of microorganism has the ability to fix atmospheric nitrogen and make the nodules in the plant roots. Many of the endangered plant species like *Meizotropis pellita* are nodulated by such type of bacteria (Pankaj et al. 2014). Beneficial soil microbes (BSMs) are identified as suitable candidates that may help in the sustainable management of the environment. These microorganisms possess several mechanisms that can be exploited at the commercial level in developing microbial biotechnology for solving the key environmental issues. Beneficial microbe-based products currently used in agroecosystems have shown remarkable success (Mishra et al. 2014). Agriculture is facing severe challenges of land degradation, lesser productivity, and susceptibility toward abiotic and biotic stresses. Sustainability in the agricultural sector is proving a formidable task because the current trend involves excessive use of chemical fertilizers and pesticides for increasing the agricultural production using pesticides in different combinations (single/mixture of pesticides). After reaching the soil, these pesticides contaminate groundwater, which also affect food chain. Reports indicated the presence of residual pesticides in various vegetables, crops, and other food products. These pesticides affect the soil microbial population as well as soil fertility parameters. Biodegradation of soil residual pesticides is possible with both bacteria and fungi. Biodegradation of cypermethrin, endosulfan, fipronil, sulfosulfuron, and imidacloprid is confirmed in laboratory through bacterial strains (*Bacillus*, *Pseudomonas*, *Xanthomonas*, *Arthrobacter*) (Negi et al. 2014, 2016; Pankaj et al. 2016a, b). These studies showed there is urgent need of compilation of all information in current aspect. From the last two decades, lots of data are generated, which can be used directly from databases and can be used for valuable analysis of the soil microorganisms. Recent omics technology is useful for the interpretation of soil microorganisms (Fig. 10.1).

10.3 Different Aspects of Bioinformatics to Study Soil Microorganism

Metagenomic approaches are now commonly used in microbial ecology to study microbial communities in more detail, including many strains that cannot be cultivated in the laboratory. Bioinformatic analysis makes it possible to mine huge metagenomic datasets and discover general patterns that govern microbial ecosystems (Hiraoka et al. 2016). Microbial bioinformatics aims to discover new biological concepts and laws based on large-scale availed data and is now expected to accelerate discovery in unexamined areas of the microbial sciences. In microbial ecology, research has been hindered because the majority of environmental microbes cannot grow up in defined media. A large number of studies across diverse natural

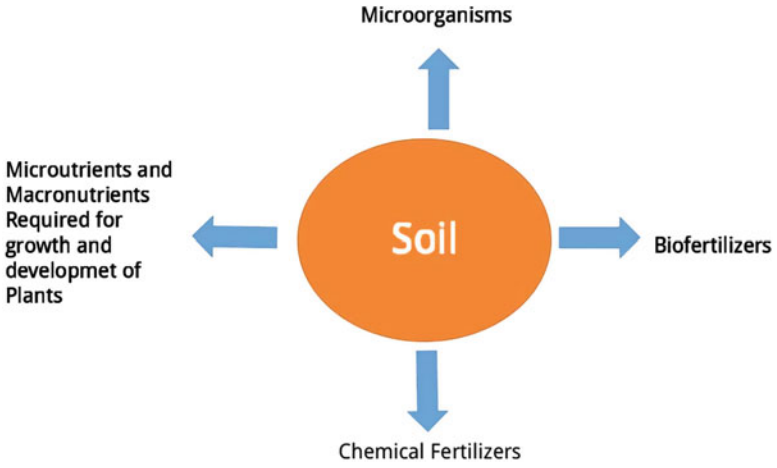


Fig. 10.1 Soil and its surrounding environment

environments have identified many microbial groups with no axenic culture. In order to overcome these fundamental difficulties, culture-independent approaches, including DNA hybridization, DNA cloning, and RT-PCR, have been used to detect specific members and/or functional genes in microbial communities. High-throughput sequencing technologies have recently popularized shotgun metagenomic and amplicon sequencing methods, which identify members and/or functional genes at a greater scale and in more details. Their use in diverse environments has revealed the presence of extremophiles, uncovered relationships between microbes and human diseases, and characterized the nutrition systems involved in symbiosis. Even more applications of these methods are used in agriculture, food science, and pharmaceuticals and forensics. Many large-scale metagenomic projects are now generating comprehensive microbial sequence collections for different environments. Since microbial communities change as they interact with other organisms and as the environmental changes, time-series analyses have also become common. DGGE (denaturing gradient gel electrophoresis) is commonly used for microbial community analysis from different environmental samples. Several bioinformatic tools have been developed and popularized to analyze metagenomic and amplicon sequence data. Web servers, such as MG-RAST, IMG/M, EBI Metagenomics, and SILVAngs, and pipelines, such as MEGAN, QIIME, and mothur, now allow researchers to perform integrated metagenomic analyses and visualize results without command-line operations or strong computational knowledge (Hiraoka et al. 2016) (Fig. 10.2) (Tables 10.1 and 10.2).

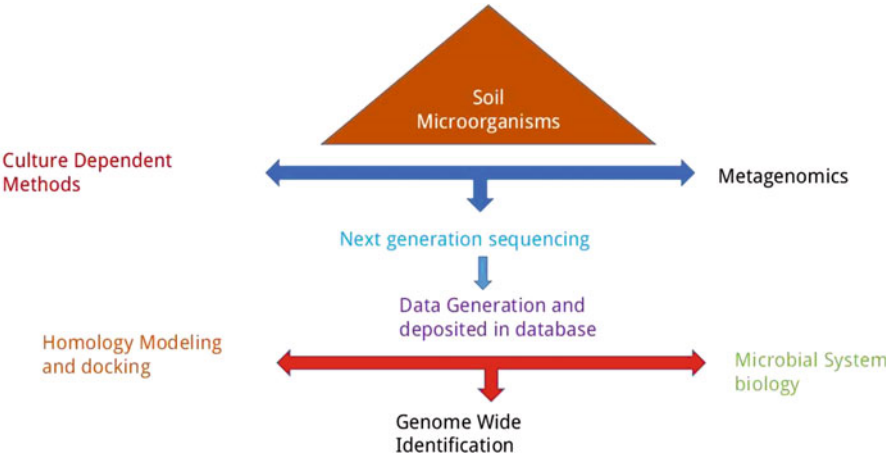


Fig. 10.2 Use of soil microorganism for different aspects of in silico studies

Table 10.1 Databases used for soil microorganisms

Database name	URL	Brief description
Aging Chart	http://www.agingchart.org/	Pathways of age-related processes
Assembly	http://www.ncbi.nlm.nih.gov/assembly	Status of whole-genome shotgun assemblies
BacWGSTdb	http://bacdb.org/BacWGSTdb/	Bacterial whole-genome sequence typing database
BIGNASim	http://mmb.irbbarcelona.org/BIGNASim/	Molecular dynamic simulations of nucleic acids
BreCAN-DB	http://brecandb.igib.res.in/	Breakpoint profiles of cancer genomes
Cancer RNA-Seq Nexus	http://syslab4.nchu.edu.tw/CRN	Transcriptome profiling in cancer cells
CauloBrowser	http://www.caulobrowser.org	Biology of *Caulobacter crescentus*
ccmGDB	http://bioinfo.mc.vanderbilt.edu/ccmGDB/	Cancer Cell Metabolism Gene DataBase
CEGA	http://cega.ezlab.org/	Conserved elements from genomic alignments
CircNet	http://circnet.mbc.nctu.edu.tw	Tissue-specific expression profiles of circular RNA
Colorectal cancer atlas	http://www.colonatlas.org	Genes and proteins of colorectal cancer cells
CRISPRz	http://research.nhgri.nih.gov/crisprz	CRISPR single-guide RNAs to zebra fish genes
CSDB	http://csdb.glycoscience.ru/database	Carbohydrate Structure Database

(continued)

Table 10.1 (continued)

Database name	URL	Brief description
DASHR	http://lisanwanglab.org/DASHR	Database of human small noncoding RNA
dbMAE	http://mae.hms.harvard.edu	Database of monoallelic gene expression
dbSUPER	http://bioinfo.au.tsinghua.edu.cn/dbsuper/	A database of super-enhancers
DESM	http://www.cbrc.kaust.edu.sa/desm	Microbial knowledge exploration systems
DIDA	http://dida.ibsquare.be	DIgenic diseases DAtabase
Digital Development database	http://cell-lineage.org	*C. elegans* development and cell differentiation
DMDD	http://dmdd.org.uk	Deciphering the mechanisms of developmental disorder
EK3D	http://www.iith.ac.in/EK3D/	Capsular polysaccharide (K antigen) structures of various *E. coli* serotypes
ENCODE DCC	http://www.encodeproject.org	ENCODE (encyclopedia of DNA elements) consortium data portal
FLOR-ID	http://www.flor-id.org/	Flowering interactive database
GEneSTATION	http://www.genestation.org	Genes in gestation: genomics of pregnancy-related tissues
GlyTouCan	https://glytoucan.org	International glycan structure repository
GreeNC	http://greenc.sciencedesigners.com/	Green non-coding: plant lncRNAs
HGTree	http://hgtree.snu.ac.kr	Horizontally transferred genes identified by tree-based methods
HPMCD	http://www.hpmcd.org/	Human Pan-microbial Communities Database
hPSCreg	http://hpscreg.eu	Human pluripotent stem cell registry
IC4R	http://ic4r.org	Information commons for rice
InsectBase	http://www.insect-genome.com/	Insect genomes and transcriptomes
InterRNA	http://mfrlab.org/interrna/	Base interactions in RNA structures
JuncDB	http://juncdb.carmelab.huji.ac.il/	Exon-exon junction database
Lnc2Cancer	http://www.bio-bigdata.com/lnc2cancer/	Human lncRNA and cancer associations
MERAV	http://merav.wi.mit.edu	Metabolic gene rapid visualizer
Metabolomics workbench	http://www.metabolomicsworkbench.org/	Metabolomic data, standards, and protocols
MitoAge	http://www.mitoage.org	Mitochondrial DNA properties and aging
MutationAligner	http://www.mutationaligner.org	Mutation hotspots in protein domains in cancer
NBDB	http://nbdb.bii.a-star.edu.sg	Nucleotide-binding protein motifs
OpenTein	http://opentein.hgc.jp/	Open teratoma investigation: images

(continued)

Table 10.1 (continued)

Database name	URL	Brief description
PCOSKB	http://pcoskb.bicnirrh.res.in/	Polycystic ovary syndrome knowledge base
PDBFlex	http://pdbflex.org	Flexibility in protein structures
PhytoPath	http://www.phytopathdb.org/	Genomics of fungal, oomycete, and bacterial phytopathogens
piRNAclusterDB	http://www.smallrnagroup-mainz.de/piRNAclusterDB.html	Clusters of piRNAs
PlanMine	http://planmine.mpi-cbg.de/	Planarian genomics
PlantDHS	http://plantdhs.org	Plant DNase I-hypersensitive sites
RBP-Var	http://www.rbp-var.biols.ac.cn/	Variation that can affect RNA-protein interactions
RMBase	http://mirlab.sysu.edu.cn/rmbase/	RNA modification database
RPFdb	http://sysbio.sysu.edu.cn/rpfdb/	Ribosome profiling database
SATPdb	http://crdd.osdd.net/raghava/satpdb/	Structurally annotated therapeutic peptides
SBR-blood	http://sbrblood.nhgri.nih.gov	Systems biology repository for hematopoietic cells
SEA	http://sea.edbc.org	Super-enhancer archive
SigMol	http://bioinfo.imtech.res.in/manojk/sigmol	Quorum-sensing signaling molecules
SIGNOR	http://signor.uniroma2.it/	Signaling network open resource
sORFs	http://www.sorfs.org	Small ORFs identified by ribosome profiling
Start2Fold	http://start2fold.eu	Hydrogen/deuterium exchange data on protein folding and stability
SureChEMBL	https://www.surechembl.org/	Chemical compounds in patent documents
SynLethDB	http://histone.sce.ntu.edu.sg/SynLethDB/	Synthetic lethality gene pairs as potential anticancer drug targets
TCGA SpliceSeq	http://projects.insilico.us.com/TCGASpliceSeq	Alternative splicing patterns in cancer cells
UET	http://mammoth.bcm.tmc.edu/uet/	Universal evolutionary trace: protein motifs important for function
WeGET	http://coexpression.cmbi.umcn.nl/	Weighted gene co-expression tool
WITHDRAWN	http://cheminfo.charite.de/withdrawn/	Withdrawn and discontinued drugs

Table 10.2 Description of databases most recently published

Database name	URL	Brief description
ANISEED	http://www.aniseed.cnrs.fr	Ascidian network for in situ expression and embryological data
BiGG models	http://bigg.ucsd.edu	Biochemically, genetically, and genomically structured metabolic network models
CPPsite	http://crdd.osdd.net/raghava/cppsite/	Validated cell-penetrating peptides
DBAASP	http://dbaasp.org	Database of antimicrobial activity and structure of peptides
DGIdb	http://dgidb.genome.wustl.edu	Drug-gene interaction database
iGNM	http://gnmdb.csb.pitt.edu/	Protein functional motions based on Gaussian network model
IID[a]	http://ophid.utoronto.ca/iid	Integrated interaction database: tissue-specific protein-protein interactions
iPPI-DB	http://www.ippidb.cdithem.fr/	Inhibitors of protein-protein interactions
KLIFS	http://klifs.vu-compmedchem.nl	Kinase-ligand interaction fingerprints and structures
MG-RAST	http://metagenomics.anl.gov/	Data portal for processing, analyzing, sharing, and disseminating metagenomic data sets
MitoCarta	http://www.broadinstitute.org/pubs/MitoCarta	Mouse and human mitochondrial proteins
MNXref/MetaNetX	http://www.metanetx.org	Genome-scale metabolic networks
MouseNet	http://www.inetbio.org/mousenet/	Functional network of mouse genes
PlantPAN	http://PlantPAN2.itps.ncku.edu.tw	Plant promoter analysis navigator
SIDER	http://sideeffects.embl.de/	Side effect resource: adverse drug reactions
sRNATarBase[a]	http://ccb1.bmi.ac.cn/srnatarbase/	sRNA-target interactions in bacteria
SugarBindDB	http://sugarbind.expasy.org	Host-pathogen interactions mediated by glycans

[a]IID and sRNATarBase have been previously listed in the NAR Database Collection as entries nos. 897 and 1832, respectively

10.4 Recent In Silico Prediction Tools

With emerging sequence data, the prediction tool became an efficient resource to study such data. The various bioinformatic tools are now available, so that gene present in the sequence, its protein, structure, and function can be predicted so that efforts can be reduced on their validation. Some of the prediction tools are discussed below:

1. **Gene prediction tools** – Genome annotation is important area to explore the existing sequence present in the databases. With existing high-throughput sequencing technologies, the current need is to annotate the existing genome for studies and practical utilities. In such studies gene prediction tools may help the researchers. Gene prediction basically utilizes the strategies of identification through transcription regulation sites, spile sites, poly A tail sites, translation start/stop sites, ORFs, and homology searching. The gene prediction is done by three different categories of software:

   (a) ***Similarity based*** – It is simplest approach of identifying input sequence similarity of gene sequences based on genes, ESTs, and protein of the other genomes. These approaches utilize local and global alignments for identification. The local alignments are done by BLAST to detect the similarity between genes, ESTs, and proteins. Global alignments use homologous protein of translated ORFs in a genomic sequence for gene prediction.

   (b) ***Ab initio based*** – Ab initio gene predictions rely on identification of gene structures rather than similarity. It utilizes gene model as a rule to identify any genes. This method uses the signal sensors and content sensors. Signal sensor denotes splice sites, poly A tail sites, and translation start/stop sites, while content sensors distinguish the exonic codons from noncoding through statistical detection algorithms. Program like GeneID and Genie can be used for gene prediction by this method (Wang et al. 2004).

   (c) ***Combined evidence based*** **–** This model combines the above model approaches, i.e., combination of gene model with alignment with known ESTs and proteins. These approaches provide best results than above two approaches. GeneSeqer is used for combined evidence approach (Brendel et al. 2004).

2. **Protein prediction tools –** Proteins are the furnished product of central dogma. With current existing sequencing technologies, the protein sequences are increasing day by day. The determination of such sequence structure and function is tedious task to the researchers. Protein prediction tools can reduce the "sequence-structure/function gap." Like gene prediction tools, protein prediction tools also use three different approaches to identify the protein which are given below:

   (a) **Homology modeling –** This is similarity-based model, principally depended upon the similarity between already existing proteins in databases, i.e., template protein and the candidate protein. This modeling is the simplest one with determination of structure by the amino acid sequences and percentage of identical residues. The homology modeling includes seven major steps, viz., template recognition and initial alignment, alignment corrections, backbone generation, loop modeling, side-chain modeling, model optimization, and model validation (Krieger et al. 2003).

   (b) ***Ab initio or de novo*** **modeling –** This modeling is done when similarly percentage is low between candidate protein and proteins present in databases (Faiza 2017). This modeling based on free energy, i.e., native protein, will

have minimum free energy in given sets of conditions. The tools useful for ab initio modeling are ROSETTA, TOUCHSTONE-II (Zhang et al. 2003), etc.

(c) **Threading –** Threading utilize the folds in both template and candidate protein for identification. It is similar to homology modeling as required specific fold through protein database. This aims to detect analogous folds in evolutionary-related proteins. The programs used for threading are THREADER, I-TASSER (Roy et al. 2010), COTH, etc.

10.5 In Silico Gel Electrophoresis

To know the gel appearance of the DNA and RNA fragments (soil microorganisms), in silico agarose gel electrophoresis is required. We can use the different combinations of the gel (0.8%, 1.0%, 1.5%) according to the desired study. Restriction digestion of the DNA fragments is possible with different restriction enzymes available with the software. SnapGene, Serial Cloner, and some others can be used on the basis of the system recommendations. These softwares are also beneficial for the study of plasmids and cloning experiment of metagenomic DNA from the soil microorganisms.

10.6 In Silico Proteomic Study of Soil Microorganisms

In silico analysis of 2D gel can be performed on the basis of isoelectric point (pI) and molecular weight of separated protein(s) as reported by Jain et al. (2010). ExPASy is a SIB Bioinformatics Resource Portal which provides access to scientific databases and software tools in proteomics, genomics, phylogeny, system biology, transcriptomics, etc. TagIdent (a software tool) is commonly used to analyze proteins that appeared as prominent spots in two-dimensional gel electrophoresis, on the basis of isoelectric point and pH (Pankaj et al. 2016a).

10.7 System Biology of Soil Microorganisms

The study of soil microorganisms is also possible with system biology methods. On the basis of the metabolites, proteins, and genes isolated and identified from soil, we can design a hypothetical system. The microbial system biology is another interesting field to depict the metabolic pathways and responsible genes using the system

biology tool (cell designer version 4.0). Many of the researchers developed the methods of system biology and their application in biological sciences. We can use the same approach using cell designer software for the identification of new system forms the soil microorganisms which might be helpful for novel microorganism's identification (Pathak et al. 2017).

10.8 Tools for DNA Barcode Analysis

Bioinformatic tools are the fastest emerging tool which enhances the efficiency of process with eradication of the labor in the analytical process. Various analytical tools are present to analyze DNA barcode or to find suitable barcode sequence for the sample. Some software and tool are also present that can perform the statistical analysis. The few are discussed below:

1. **MUSCLE** – It is a multiple protein sequence alignment software. This software is important as it aligns the query sequence with the reference dataset sequence to find out the correct taxon. It is one fast and accurate tool for the alignment (Edgar 2004).
2. **BRONX** (Barcode Recognition Obtained with Nucleotide eXposes) – Alignment sometimes doesn't consider within taxon variability which is due to haplotype and thus can lead to misidentification. Therefore BRONX is used as its analysis is based on alignment-free and character-based approach. These utilize the flanking region (pretext and postext) of query sequence and on the basis of that flanking sequence score the variable region of the query sequence, thus identifying the variability which is used for scoring (Little 2011).
3. **CBCAnalyzer** – CBC analyzer works on principle of identification of compensatory base changes. When a changed nucleotide at both paired sites remains attached and pairing is maintained, it is called CBC mutation. Single mutation present in conserved ITS2 secondary structure, indicates that the sexual intercrossing is not possible between two species which serve as a rule for the discrimination between two species (Coleman and Vacquier 2002). The CBC analyzer is a triple component-free program which contains CTTransform, CBCDetect, and CBCTree. The CTTransform transforms the input file in bracket dot bracket format file which is read by other tools like MARNA for creating multiple alignments based on primary sequence and secondary structure. This alignment information is used as input by CBDetect to find out the CBC mutations. This CBC is used as input for CBCTree and makes phylogram by detecting sexually incompatible species (Wolf et al. 2005).

4. **SPIDER (Species IDentity and Evolution in R) –** It is R package which runs on R platform utilized for DNA barcoding for various types of analysis related to species delimitation, evolution, and speciation. The barcode analysis mainly includes summary statistics, sliding window analysis, and taxonomy and evolution. Summary statistics provide information about number of individuals, length of sequence, number of haplotypes in particular species, missing data proportion, and number of species. This also estimates the barcoding gap. This R package is also equipped with graphical and tabular function to depict the results. Another important feature in this package is sliding window analysis across the sequences. It can partition the full segment into smaller frame, and further testing can be done (Brown et al. 2012).
5. **ecoPrimers –** It is an open-source software which can scan whole genome for finding the conserved markers and thus allows the selection of marker without identification. The primer pair is selected by estimation of amplification range and taxonomical discrimination capacity. This criteria optimization thus concludes whether the primer pair can be used for fungal barcoding or not (Riaz et al. 2011). Thus ecoPrimers scan be implemented in a way that allows the design of new barcodes specific to any taxonomic group.

The various other softwares and tools are TaxI (Steinke et al. 2005), 4 SALE (Seibel et al. 2006), CLOTU (Kumar et al. 2011), etc. These all software tools and program made barcoding research findings more predictive, reliable, and authentic thus simplifying taxon identification process. With the help of these tools, we can predict the novel view of soil microbial communities.

10.9 Artificial Neural Network (ANN)

Artificial neural network is biologically inspired (structure, processing method, learning ability) computer program that works similarly as human brain functions in different aspects. It is formed from hundreds of single units, artificial neurons, or processing elements (PE), connected with coefficients (weights), which constitute the neural structure and are organized in layers. The power of neural computations comes from connecting neurons in a network. Each PE has weighted inputs, transfer function, and one output. The behavior of a neural network is determined by the transfer functions of its neurons, by the learning rule, and by the architecture itself (Kustrin and Beresford 2000). Microbial communities vary from one place to another, but prediction of their biomass is a difficult challenge due to large number of variables that affect communities. ANN models are useful for the prediction of such types of difference between different environments (Santos et al. 2014).

10.10 Conclusion

Using in silico study, we can perform the more accurate analysis without expensing the chemicals in wet laboratory. There are many databases dealing with the omics technology, which can be fruitful for the study of microbial behavior. DNA barcoding can also be used by taxonomist for identification of microorganisms. Bioinformatic-based tools also deal with metagenomic-based data. On the basis of these databases and softwares, we can analyze phylogenetic, genomic, transcriptomic, and proteomic data.

Acknowledgment Authors P.B and A.B are thankful for all the researchers for their contribution. Their contribution cited in reference section.

References

Brendel V, Xing L, Zhu W (2004) Gene structure prediction from consensus spliced alignment of multiple ESTs matching the same genomic locus. Bioinformatics 20:1157–1169. https://doi.org/10.1093/bioinformatics/bth058

Brown SDJ, Collins RA, Boyer S et al (2012) Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. Mol Ecol Resour 12:562–565. https://doi.org/10.1111/j.1755-0998.2011.03108.x

Coleman AW, Vacquier VD (2002) Exploring the phylogenetic utility of ITS sequences for animals: a test case for abalone (Haliotis). J Mol Evol 54:246–257. https://doi.org/10.1007/s00239-001-0006-0

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340

Faiza M (2017) Ab-initio prediction of protein structure: an introduction. Bioinforma Rev 2:4

Hiraoka S, Yang C-C, Iwasaki W (2016) Metagenomics and bioinformatics in microbial ecology: current status and beyond. Microbes Environ 31:204–212. https://doi.org/10.1264/jsme2.ME16024

Jain A, Bacolla A, Chakraborty P et al (2010) Human DHX9 helicase unwinds triple-helical DNA structures. Biochemistry 49:6992–6999. https://doi.org/10.1021/bi100795m

Kress WJ, Erickson DL (2012) DNA barcodes: methods and protocols. In: Methods in molecular biology. Clifton, pp 3–8

Krieger E, Nabuurs SB, Vriend G (2003) Homology modeling. In: Bourne P, Weissig H (eds) Structural bioinformatics. Wiley-Liss, Hoboken, pp 507–521

Kumar S, Carlsen T, Mevik BH et al (2011) CLOTU: an online pipeline for processing and clustering of 454 amplicon reads into OTUs followed by taxonomic annotation. BMC Bioinforma 12:182. https://doi.org/10.1186/1471-2105-12-182

Kustrin AS, Beresford R (2000) Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J Pharm Biomed Anal 22:717–727

Little DP (2011) DNA barcode sequence identification incorporating taxonomic hierarchy and within taxon variability. PLoS One 6:e20552. https://doi.org/10.1371/journal.pone.0020552

Mishra J, Prakash J, Arora NK (2014) Role of beneficial soil microbes in sustainable agriculture and environmental management. Clim Chang Environ Sustain 4:137–149

Negi G, Pankaj, Srivastava A, Sharma A (2014) In situ biodegradation of endosulfan, imidacloprid, and carbendazim using indigenous bacterial cultures of agriculture fields of Uttarakhand, India. Int J Bioeng Life Sci 8:973–981

Negi G, Srivastava A, Sharma A (2016) Optimization of endosulfan biodegradation using indigenous bacterial isolate bacillus Aryabhatti through response surface methodology. J Ind Pollut Control 32:638–644

Nielsen UN, Wall DH (2013) The future of soil invertebrate communities in polar regions: different climate change responses in the Arctic and Antarctic? Ecol Lett 16:409–419. https://doi.org/10.1111/ele.12058

O'Donnell AG, Young IM, Rushton SP et al (2007) Visualization, modelling and prediction in soil microbiology. Nat Rev Microbiol 5:689–699. https://doi.org/10.1038/nrmicro1714

Ogram A (2000) Soil molecular microbial ecology at age 20: methodological challenges for the future. Soil Biol Biochem 32:1499–1504. https://doi.org/10.1016/S0038-0717(00)00088-2

Orgiazzi A, Dunbar MB, Panagos P et al (2015) Soil biodiversity and DNA barcodes: opportunities and challenges. Soil Biol Biochem 80:244–250. https://doi.org/10.1016/J.SOILBIO.2014.10.014

Pankaj, Nailwal TK, Singh L, Panwar A (2014) Isolation and characterization of rhizobial isolates from the rhizospheric soil of an endangered plant meizotropis pellita. Asian J Microbiol Biotechnol Environ Sci Pap 16:301–306

Pankaj, Negi G, Gangola S et al (2016a) Differential expression and characterization of cypermethrin-degrading potential proteins in *Bacillus thuringiensis* strain, SG4. 3 Biotech 6:225. https://doi.org/10.1007/s13205-016-0541-4

Pankaj, Sharma A, Gangola S et al (2016b) Novel pathway of cypermethrin biodegradation in a Bacillus sp. strain SG2 isolated from cypermethrin-contaminated agriculture field. 3 Biotech 6:45. https://doi.org/10.1007/s13205-016-0372-3

Pathak RK, Baunthiyal M, Pandey N et al (2017) Modeling of the jasmonate signaling pathway in *Arabidopsis thaliana* with respect to pathophysiology of alternaria blight in brassica. Sci Rep 7:16790. https://doi.org/10.1038/s41598-017-16884-3

Probandt D, Eickhorst T, Ellrott A et al (2017) Microbial life on a sand grain: from bulk sediment to single grains. ISME J. https://doi.org/10.1038/ismej.2017.197

Riaz T, Shehzad W, Viari A et al (2011) ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. Nucleic Acids Res 39:e145. https://doi.org/10.1093/nar/gkr732

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725–738. https://doi.org/10.1038/nprot.2010.5

Santos EC, Armas ED, Crowley D, Lambais MR (2014) Artificial neural network modeling of microbial community structures in the Atlantic Forest of Brazil. Soil Biol Biochem 69:101–109. https://doi.org/10.1016/J.SOILBIO.2013.10.049

Seibel PN, Müller T, Dandekar T et al (2006) 4SALE – a tool for synchronous RNA sequence and secondary structure alignment and editing. BMC Bioinforma 7:498. https://doi.org/10.1186/1471-2105-7-498

Steinke D, Vences M, Salzburger W, Meyer A (2005) TaxI: a software tool for DNA barcoding using distance methods. Philos Trans R Soc Lond Ser B Biol Sci 360:1975–1980. https://doi.org/10.1098/rstb.2005.1729

Wang Z, Chen Y, Li Y (2004) A brief review of computational gene prediction methods. doi:https://doi.org/10.1016/S1672-0229(04)02028-5

Whitman WB, Coleman DC, Wiebe WJ (1998) Prokaryotes: the unseen majority. Proc Natl Acad Sci 95:6578–6583

Wolf M, Friedrich J, Dandekar T, Müller T (2005) CBCAnalyzer: inferring phylogenies based on compensatory base changes in RNA secondary structures. In Silico Biol 5:291–294

Zhang Y, Kolinski A, Skolnick J (2003) TOUCHSTONE II: a new approach to ab initio protein structure prediction. Biophys J 85:1145–1164. https://doi.org/10.1016/S0006-3495(03)74551-2

Chapter 11
Sustainable Agriculture: Role of Metagenomics and Metabolomics in Exploring the Soil Microbiota

Neeraj Gupta, Siddharth Vats, and Prachi Bhargava

Abstract Land-use change and agricultural management have important effects on soil microbiome, as they change the physical and chemical properties of soil. In addition, agricultural management (e.g., tillage, pesticide, and fertilizer applications) directly affects soil biodiversity by altering the physical and chemical properties of soil. Soil metagenomics, which comprises of isolation of soil DNA and the production and screening of clone libraries, can provide a cultivation-independent assessment of the largely untapped genetic reservoir of soil microbial communities. This approach has already led to the identification of novel genes and biomolecules. Metabolism of different microorganisms that change in response to different environmental conditions can be studied by analysis of metabolic footprinting. The concentration of extracellular metabolites varies depending on the factors including temperature, pH, and the concentration of nutrients. These factors affect the uptake and secretion of metabolites from soil. This chapter describes how these novel tools have helped to explore the microbiota of the soil and how these can be used as a guiding line to innovate new agricultural norms for sustainable environment.

Keywords Sustainable agriculture · Bacterial diversity · Shotgun sequencing · Metagenomics · Metabolomics

11.1 Sustainable Agriculture

Pedosphere is the most complex environment consisting of soil which appears to be a major reservoir of microbial genetic diversity. Sustainable agriculture is an apt answer to the challenges the world is facing today in terms of deterioration of quality and quantity of all the natural resources used to grow crops with the advent of modern agriculture. The complex relationship between agriculture, environmental

N. Gupta · S. Vats · P. Bhargava (✉)
Institute of Biosciences and Technology, Shri Ramswaroop Memorial University, Barabanki, Uttar Pradesh, India
e-mail: prachi.bio@srmu.ac.in

D. K. Choudhary et al. (eds.), *In Silico Approach for Sustainable Agriculture*,
https://doi.org/10.1007/978-981-13-0347-0_11

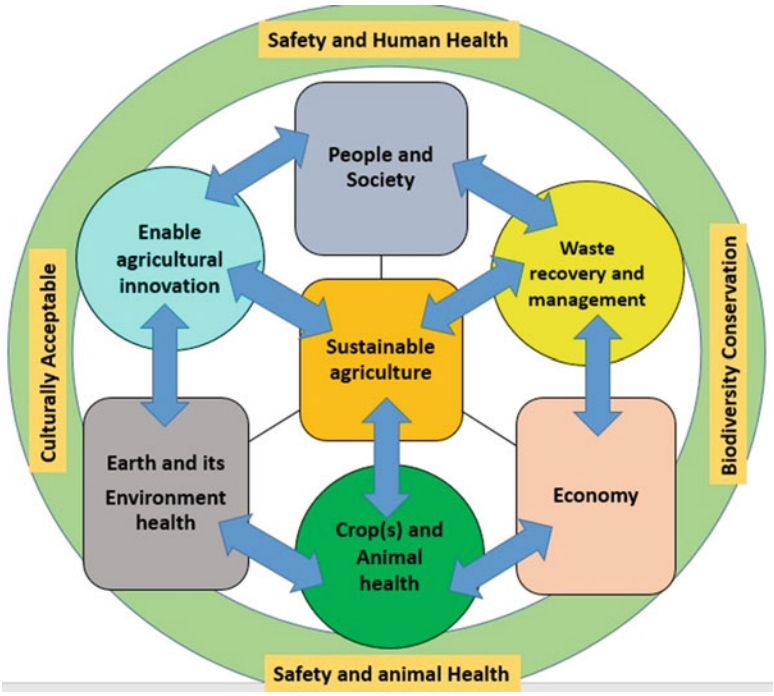


Fig. 11.1 Integration of sustainable agriculture for sustainable society

sustainability, and social systems suggests that the holistic development results from the complex interaction of a multitude of factors (Fig. 11.1). Dependence on chemicals for further agricultural needs will result in future loss in soil physical condition, feasibility of water pollution, and calculated burden on the fiscal system. Inaugurating an ecological friendly parallel mechanism on earth is of vital importance.

11.2 Challenges to Sustainable Agriculture

Almost all the countries in the world are more or less, directly or indirectly dependent on agriculture for their economy. The use of modern techniques and intensive farming ensures a good yield but simultaneously harms the environmental sustainability. Urbanization is the biggest threat to soil biodiversity which is depleting the enormous pool of microflora and fauna inhabiting the soil. Figure 11.2 summarizes the major challenges the current agriculture faces in keeping a fine balance between maintaining the sustainability and increasing fiscal growth rate.

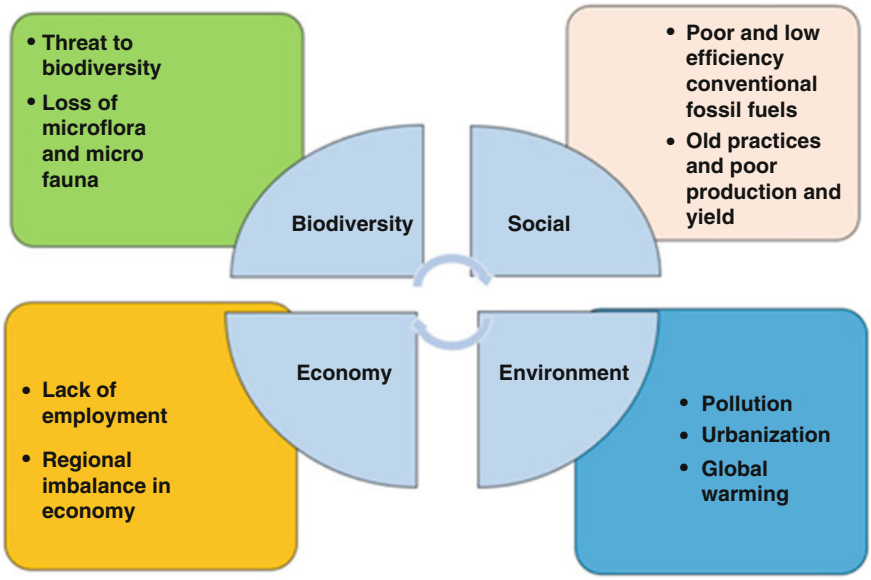


Fig. 11.2 Challenges to sustainable agriculture

11.3 Metagenomics

Our environment is a major reservoir of microbial genetic diversity (Robe et al. 2003). The complexity of microbial diversity results from multiple interacting parameters, which include pH, water content, soil structure, climatic variations, and biotic activity.

Most of the soil microorganisms seem extremely well adapted to their environment, however, cannot be cultured under the usual laboratory conditions. Current estimates indicate that more than 99% of the microorganisms present in many natural environments are not readily culturable and therefore not accessible for biotechnology or basic research (Schloss and Handelsman 2003). Microbial communities are important in the functioning of all ecosystems, but the unculturable microorganisms and their role in natural ecosystems are unclear (Tyson et al. 2004). Microorganisms vary in the ways in which they forage for food, transduce energy, contend with competitors, and associate with allies. But the variations that we know are only the tip of the microbial iceberg.

Metagenomics can unlock the massive uncultured microbial diversity present in the environment for new molecules for therapeutic, biotechnological, and sustainable agricultural applications. This technique entails extraction of DNA from a community so that all of the genomes of organisms in the community are pooled. The purified soil metagenomic DNA can be cloned into several different plasmid vectors. In general, high-molecular-weight DNAs cloned into fosmid, cosmid, or bacterial artificial chromosome (BAC) vectors are introduced into a surrogate host

bacterium, such as *Escherichia coli* (Rondon et al. 2000). Especially, the fosmid-based preparation of a metagenomic library is the most frequently used strategy, due to their high- cloning efficiency, the improved stability in *E. coli*, and the feasibility to construct a medium-sized (40 kb) insert DNA library (Kim et al. 1992). Considering the low frequency of finding genes for novel bioactivities from a soil metagenomic library (Chung et al. 2008), the cloning efficiency is a crucial factor to construct a large clone member library. One of the most critical steps for successful library construction is the isolation of pure high-molecular-weight DNA, which is suitable for cloning into the vectors (Delmont et al. 2011). These genomes are usually fragmented and cloned into an organism that can be cultured to create "metagenomic libraries," and these libraries are then subjected to analysis based on DNA sequence following sequence-driven approach or on the functions conferred to the host by the environmental DNA.

11.4 Experimental Strategies Involved in Metagenomics

11.4.1 Soil DNA Extraction: Direct and Indirect

Metagenomic DNA from a microbial population can be extracted by direct and indirect extraction methods. The ideal methodology should include unbiased lysis of cells and extraction of their nucleic acids. Direct DNA isolation is based on cell lysis within the sample matrix and subsequent separation of DNA from the matrix and cell debris (Ogram et al. 1987). It is designed to release the DNA by breaking the cell wall and membranes of the microorganisms. In this method, any extracellular DNA is first separated from the environmental sample by treating it with an alkaline buffer. The cells in the matrix are then subjected to direct mechanical (e.g., bead beating) lysis followed by extraction of DNA released from these cells. DNA recovered by centrifugation is then concentrated and purified before cloning.

In contrast, the indirect method involves the separation of cells from the soil matrix followed by cell lysis and DNA extraction. The recovered cells are subjected to cell lysis (chemical and enzymatic) followed by DNA extraction and purification (Holben et al. 1988). Cell lysis is a critical and time-consuming step in soil metagenomic DNA extraction. Although the indirect extraction method prevents contamination from nonbacterial DNA that may be present in the sample. Chemical or enzymatic lysis is relatively gentle, and they often discriminate against particular cell types and do not completely penetrate soil or sediment samples. Mechanical disruption gives more uniform cell disruption and disperses soil or sediment samples to allow good penetration of the lysis buffer. Therefore, mechanical treatment is more effective and less selective than chemical lysis. The mechanical disruption methods include thermal shocks, bead-mill homogenization, bead beating, microwave heating, and ultrasonication. Thermal shock consists of repeated freezing and thawing of the sample suspensions. The number of freeze–thaw cycles and the incubation time and temperatures can be varied. Thermal shock is less violent than

other mechanical treatments such as microwave heating, ultrasonication, and bead beating (Cullen and Hirsch 1998). Ultrasonication treatment efficiently releases the bacterial cells bound to the soil aggregates. Power and duration of sonication treatment are optimized based on the lysis efficiency and the shearing of DNA.

Direct extraction methods provide high yield of lower size DNA fragments, whereas indirect methods provide low yield of higher size DNA fragments. Both methods have distinct advantages and limitations, and the choice should be based on the intended downstream application and the objective of the study.

While the microbial diversities of various microbial habitats are being actively investigated by taking advantage of next-generation sequencing technology to analyze a large number of 16S rRNA gene amplicon sequences derived from a variety of soils (Delmont et al. 2011), the advances in bioprospecting metagenomics are relatively slow. The physical and chemical structure of each microbial community affects the quality, size, and amount of microbial DNA that can be extracted. When isolating DNA from the environment for the construction of large insert met genomic libraries, three problems have to be solved. (i) The DNA should be extracted from as broad range of microorganisms as possible so that it is a representative of the original microbial population. (ii) Shearing of DNA has to be avoided during the extraction procedure because high-molecular-weight DNA is required for suitable community analysis; smaller DNA fragments may lead to formation of chimeric products. (iii) The DNA must be free from contaminating substances which interfere with downstream DNA processing such as restriction and ligation. Inorganic soil components, such as negatively and positively charged clay particles, and biochemical contaminants, such as humic acids and DNases, make DNA extraction from soils, and subsequent manipulation, challenging. The process for removing contaminants determines both the clonability and the size of the DNA because many of the processes that effectively remove contaminants that inhibit cloning also shear the DNA (Schmeisser et al. 2007).

11.4.2 Construction of Metagenomic Library

Purified metagenomic DNA requires the construction of metagenomic DNA library and its efficient screening to discover the novel metabolites (Daniel 2004). With the help of metagenomic libraries, it is now possible to explore diversity of microbes in uncultured system and form the basis of genomic studies to link functional and phylogenetic relationship of microbes and environment. The classical method of metagenomic library construction involves insertion of small sequences of less than 10 kb into a standard sequencing vector, but small insert libraries do not allow the detection of large gene clusters or operons, and a large number of clones would have to be screened (Li et al. 2005; Walter et al. 2005). Therefore the soil metagenomic DNA would be cloned into several different plasmid vectors. In general, high-molecular-weight DNAs cloned into fosmid, cosmid, or bacterial artificial

chromosome (BAC) vectors, using insert sizes of approximately 40–200 kb, are introduced into a surrogate host bacterium, such as *Escherichia coli* (Rondon et al. 2000).

11.4.3 Analysis of Metagenomic Library

Metagenomic libraries are analyzed for information about novel enzymes by two types of analysis: a sequence-driven approach, in which libraries are initially screened for particular DNA sequences, and a function-driven approach, in which metagenomic libraries are initially screened for an expressed trait.

11.4.3.1 Sequence Driven Analysis

Sequence-based metagenomics relies on the use of conserved DNA sequences to collect genomic information from microbes without culturing them. It is used to design hybridization probes or PCR primers to screen metagenomic libraries for clones that contain sequences of interest. With this approach marker genes are screened using DNA probe or PCR primers designed from DNA sequences of already known genes. This involves the complete sequencing of clones containing phylogenetic anchors, such as the 16S rRNA gene and the archaeal DNA repair gene *rad*A (Béjà et al. 2002; Quaiser et al. 2003), which indicates the taxonomic group and functional information about the organisms from which these clones were derived.

The sequence conservation of regions of phylogenetic anchors facilitates their isolation without prior knowledge of the full gene sequence. Random sequencing can be another alternative where a gene of interest is identified, followed by screening of phylogenetic anchors in the flanking DNA. The significance of sequence-based screening is its independence from the expression of cloned genes by foreign hosts because few classes of genes contain sufficiently conserved regions to facilitate their identification by sequence instead of function. The genes encoding polyketide synthases (PKSs) and peptide synthetases, which contribute to synthesis of complex antibiotics, are the best examples. The PKSs are modular enzymes with repeating domains containing divergent regions that provide the variation in chemical structures of the products. These regions are flanked by highly conserved regions, which have provided the basis for designing probes to identify PKS genes among metagenomic clones (Courtois et al. 2003). On the other hand, using the archaeal small subunit rDNA as phylogenetic marker, DeLong and coworkers (Stein et al. 1996) identified 38.5 kbp recombinant fosmid clones, of which random shotgun sequencing revealed several genes, including translation EF2, having similarities with the archaeal homolog. Afterward, they identified bacteriorhodopsin-like gene in an insert containing a 16S rRNA gene that affiliated with the γ-*Proteobacteria* (Béjà et al. 2001).

Sequence-driven analysis of a metagenome generates a large dataset which should be analyzed by support of a combination of appropriate bioinformatics, such as comparative sequence analysis.

Since the soil microbial community is the most complicated one with the highest microbial diversity compared with any other microbial community, the analysis of the soil metagenome by the sequencing-based approach is still a huge challenge. To overcome this difficulty, comparative sequence analysis with the large set of soil metagenome sequences will be a new opportunity to search for microbial enzymes and novel bioactivities in the future (Teeling and Glöckner 2012). Another prokaryotic gene-finding program is a software tool, MetaGene, which utilizes besides other various measures two sets of codon frequency interpolations, one for bacteria and one for archaea, estimated by the guanine–cytosine (GC) content of a given sequence. Almost all annotated genes (96% of known genes and 92% hypothetical genes) and in addition about 0.4 million novel genes were predicted by MetaGene from metagenomic sequences of the Sargasso Sea dataset (Noguchi et al. 2006). As patterns emerge in the environmental sequences, sequence-based methods will be increasingly more informative about microbial communities.

A number of technologies are currently under development using microarrays. For instance, the DOTUR software was developed and used to determine whether a library contains sufficient genes for it to be considered representative of the diversity in the original microbial community (Schloss and Handelsman 2005).

11.4.3.2 Function Driven Analysis

Function-based analysis is a widely used technique which involves initial screening of clones expressing a desired trait, followed by characterization of the active clones by sequence and biochemical analysis. Through function-based analysis, identification of new enzymes, antibiotics, or other reagents in libraries from diverse environments is quite easy. An approach includes first heterologous expression to express clones of desired function and second is selection of only the clone expressing the desired function to finding rare clones so that only desired clones grow and others do not. Sequencing of the highly conserved genes such as one encoding 16S rRNA or flanking DNA of some clones reveals a gene or a group of genes that can be used to infer the phylogenetic relationship of the organism from which the DNA in the clone was isolated. On the other hand, DNA polymerase or RecA-codon can be used in the absence of a conserved gene, and sequence alignment of gene clusters with genes in the databases can be used to make phylogenetic inferences.

Mostly *Escherichia coli* is preferred as host strain, although there are other hosts used in metagenomic studies. But many genes have different expression requirements, which cannot be expressed in *E. coli.* So to extend the range of functional screening, *Streptomyces lividans* and *Pseudomonas putida* have been developed as alternative hosts (Martinez et al. 2004). *S. lividans* is a particularly useful host for

functional screening and detection of a range of other novel metabolites of soil metagenomic libraries (Courtois et al. 2003; Wang et al. 2000).

Functional screening quickly identifies clones that have medicinal, agricultural, or industrial uses by focusing on natural products or proteins that have useful activities and phylogeny of the organisms having these genes. Besides their useful approach, it also has some limitations: (i) it requires expression of the function of interest in the host cell and clustering of all of the genes required for the function and (ii) it depends on the availability of an assay for the function of interest that can be performed efficiently on vast libraries, because the frequency of active clones is quite low.

Functional analysis of metagenomic libraries has identified novel degradative enzymes (Knietsch et al. 2003; Voget et al. 2003), antibiotics (Courtois et al. 2003; MacNeil et al. 2001; Wang et al. 2000; Venter et al. 2004), antibiotic resistance genes (Riesenfeld et al. 2004), lipases (Rondon et al. 2000; Henne et al. 2000), chitinases (Cottrell et al. 1999), membrane proteins (Majerník et al. 2001), genes encoding enzymes for the metabolism of 4-hydroxybutyrate (Henne et al. 1999), and genes encoding the biotin synthetic pathways (Entcheva et al. 2001).

Instead of in vitro systems, scientists have developed a high-throughput method metabolite-related expression (METREX) in which the activity sensor is in the same cell as the metagenomic DNA. Using this method, Williamson (Tyson et al. 2004) identified clones producing a quorum-sensing inducer for expression of green fluorescent protein (GFP) (Williamson et al. 2005). Additional innovative screening approaches such as substrate-induced gene expression screening (SIGEX) have facilitated the cloning of catabolic operons, the expression of which is generally induced by relevant substrates or controlled by promoters present near the catabolic genes and potentially involved in benzoate and catechol degradation among others (Uchiyama et al. 2005). This method involved an operon trap gfp-expression vector for generation of the metagenomic clones. The clones are then incubated in the presence of the target substrate and positive clones identified by fluorescence-activated cell sorting (FACS). SIGEX is a highly successful technology for screening of large metagenomic libraries. Pre-amplification inverse-PCR (PAIPCR), SNP (single-nucleotide polymorphism) analysis, and metagenomic DNA shuffling are recently developed novel methods to isolate new biocatalysts (Spencer et al. 2003).

11.5 Metabolomics

Like genomics and proteomics, metabolomics is the study of complete set of metabolites that are present within a biological sample. Metabolomics is a relatively new approach aimed to improve genetically modified plants, understanding the metabolic networks and metabolites present inside and outside the cells. Analytical

techniques used for metabolomics analysis include mass spectrometry (MS), nuclear magnetic resonance (NMR), and HPLC.

11.6 Experimental Strategies Involved in Metabolomics

11.6.1 Sample Preparation and Separations of Extracellular Metabolites

Metabolomic sample preparation is the first and the most important step during analysis. Extracellular metabolites are separated by centrifugation or fast filtration. The supernatant containing extracellular metabolites are stored at a low temperature (−20 °C). At low temperature, some metabolites can lose their structure, characteristics, and therefore their function. To resist the properties of metabolites, they must be kept under vacuum (Villas-Bôas and Bruheim 2007). Extracellular metabolites may be liquid or volatile. On the basis of state of metabolites, the preparation of sample also varies.

Extracellular metabolites are present in the microbial culture media containing high concentration of salts, proteins, lipids, sugars, and water. This can cause difficulty with the operational efficiency of methodical equipments as well as in quantitative estimation of those metabolites that were present in low concentration in samples (Alvarez-Sanchez et al. 2010). This problem needs to be identified and resolved to estimate the accurate concentration of metabolomes by using liquid-liquid separation using immiscible solvent, column chromatography in which resin is used to bind with metabolites, and solid-phase extraction (SPE), to trap the metabolomes (Villas-Bôas and Bruheim 2007; Pinu and Villas-Boas 2017). SPE is a process of sample preparation in which metabolites that are dissolved or suspended in a liquid mixture are separated on the basis of their physical and chemical properties from other molecules present in the sample mixture. Sorbent materials used in SPE include alkylated silica, ion exchange materials, etc.

The stationary phase includes normal phase SPE procedure, reversed-phase SPE, and ion exchange SPE. In normal phase, the extraction of in vivo nonpolar analytes is done by using a nonpolar matrix and polar stationary phase. Reversed-phase SPE has a nonpolar stationary phase and mobile phase with polar or quite polar matrix. Charged metabolites are isolated in ion-exchange phase. Metabolites in solution are also separated by solid-phase microextraction (SPME). SPME is a solventless sample preparation and extraction technique.

Volatile metabolites are separated by headspace analysis or headspace coupled to solid-phase mircoextraction. Volatile or gaseous extracellular metabolites in samples give valuable information of microorganisms present in soil and their metabolic pathways. Gaseous metabolites are very difficult to extract because of their ultra-low concentrations and high diffusion rates. Therefore, metabolites must be concentrated

by using vacuum-drying and freeze-drying before analysis by analytical instruments to improve their detection (Villas-Bôas and Bruheim 2007).

11.6.2 Techniques Used for Analysis of Metabolomics Profile

Various techniques are used for metabolomic analysis such as high-performance liquid chromatography (HPLC) with UV detection, Fourier-transformed infrared (FT-IR) spectroscopy, nuclear magnetic resonance (NMR), and mass spectrometry in combination with chromatography such as gas chromatography–mass spectrometry (GC-MS) and liquid chromatography–mass spectrometry (LC-MS) are utilized to detect a wide range of metabolites with relatively high specificity and reproducibility (Defernez et al. 2004; Gidman et al. 2003; Wilson et al. 2005; Lenz et al. 2005). HPLC methods are more sensitive and reliable. Using these techniques, many series of metabolites have been analyzed in biological samples. HPLC is more sensitive and robust analytical method when it is combined with MS for the metabolomic study, and specific softwares are available for identification of metabolites.

LC-MS or LC-TOF MS (time-of-flight mass spectrometry) in association with electrospray ionization for residue analysis is used for the study of many compounds in plants (Valverde et al. 2010). The characterization and variability of metabolites in different parts of plants in different geographical is done by LC-MS (Hoffmann et al. 2017). GC-MS is a significant analytical tool for identification of metabolites. GC-MS combined with genetic algorithm (GA-DFA) analysis is useful for identification of specific metabolites that are biologically important (Shuman et al. 2011). Metabolomic analysis by GC-MS shows potential metabolites and pathways involved in lead and cadmium stress response of radish roots (Wang et al. 2015). Capillary electrophoresis–mass spectrometry (CE-MS) has been used to analyze sugar nucleotides in bacterial metabolomics (Soo et al. 2004). Various NMR techniques have been employed to study metabolomics. There have been various reports on the different applications of NMR in plants metabolomics (Schripsema 2010).

11.7 Applications of Metagenomics and Metabolomics in Sustainable Agriculture

11.7.1 In Assessment of Genetic Variation

The microbiota comprising of the soil-inhabiting microflora and microfauna play an important role in sustainable agriculture which in return depends upon integrated nutrient management (INM) and soil biodiversity system (SBS). The wide application of microorganisms in sustainable agriculture is due to the genetic reliance of

plants on the beneficial functions given by symbiotic cohabitants (Noble and Ruaysoongnern 2010). Metagenomics help in the prediction of the microbiota community structure, and, therefore, it can be used in addressing fundamental scientific questions related to agriculturally important microbes.

This approach has been successfully explored for the assessment of the microbes belonging to the rhizosphere of many crops by targeting genes which are present in microbes, but their products are beneficial for crops like the gene involved in nitrogen fixation (*nif*H) and cold shock proteins (csp) (Suyal et al. 2015). Generally the community structure and diversity of such beneficial microbes as well as of those involved in plant pathogen interactions can be explored to provide the backbone for further studies.

Simultaneously metabolic profiles can be used as signatures for assessing the genetic variation among different cultivars or species or of the same genotypes at different growth stages and different environments including the biotic and abiotic stresses. Metabolite profiles in combination with data generated from different omics experiments profiling the transcriptome, proteome, and genetic variations like single-nucleotide polymorphisms can be used to map the loci underlying various metabolites and to link these loci to crop phenotypes. Moreover, both these tools can also be explored for reshaping the composition of rhizospheric microbial population and to readdress microbial activity, which can be referred to as "rhizosphere engineering."

11.7.2 In Paleogenomics

As we know that soil DNA does not limit itself to any one species of hereditary material. Ancient DNA of plants and animals is not only damaged and fragmented but also mixed with the genomes of the abundant diverse microbes. PCR-amplified mitochondrial sequences have been used to resolve phylogenetic relationships between extinct and modern animals (Hofreiter et al. 2004). A metagenomic and metabolomic approach along with high-throughput sequencing provides a means to access the nuclear genomes of extinct organisms (Kim et al. 2006). These approaches open up the possibility for genome projects targeting extinct species and could revolutionize paleobiology (Li and Liu 2006).

11.7.3 In Exploring Microbes Which Affect Biogeochemical Cycles

Biogeochemical cycles are the pillars on which environmental sustainability stands upon. Microbes play a pivotal role in the functioning of biogeochemical cycles that are responsible for the environment of soils and oceans. Metagenomics has

successfully explored the genome assembly of certain uncultured microbes (*Kuenenia stuttgartiensis*) which have associations with anaerobic ammonium oxidation (Strous et al. 2006). Tyson and co-workers (2004) extracted DNA directly from the biofilm and obtained nearly complete genome sequences of *Leptospirillum* group II and *Ferroplasma* type II and partial genome sequences for *Leptospirillum* group III, *Ferroplasma* type I, and G-plasma. Using combined shotgun and MS-based proteomics with community genomic analysis, high expression has been reported of proteins related to oxidative stress along with a novel protein cytochrome, which is an essential component of iron oxidation and acid mine drainage formation (Ram et al. 2005).

11.7.4 In Searching Novel Biocatalysts

Biocatalysts are the essential class of proteins, playing vital roles in all biological organisms. They play an important role in metabolic processes. Identification of substrate specificity of novel enzymes is remarkable to understand the cell's metabolic process, disclose attractive drug targets, and develop new enzymes for industrial applications (Schmid et al. 2001). Novel biocatalysts can be searched by using different metagenomic and metabolomic tools (Ferrer et al. 2005; Piel 2011; Lorenz et al. 2002). Metagenomic and metabolomic methods have been developed for identification of novel enzymes (Saito et al. 2006). Metagenomic sequencing is a powerful method for discovery of gene function and novel enzymes (Xing et al. 2012). Through metagenomics various enzymes like lipases, amylases, cellulases, proteases, and xylanases have been identified (Nazir 2016). Various techniques including in silico, X-ray crystallography, activity-based protein profiling, activity-based metabolomic profiling, and capillary electrophoresis–electrospray ionization–mass spectrometry (CE-ESI-MS) find application in the detection of novel enzymes (Fig. 11.3). They rely on specific changes in the metabolite composition which can directly suggest the presence of an enzymatic activity (Prosser et al. 2014; Saito et al. 2006).

11.7.5 In Exploring Novel Sources of Antibiotics

Functional metagenomics is a very powerful tool for constructing metagenomic library for identification of novel antibiotics, which are among the most successful therapeutic compounds. Two important antibiotic compounds, namely, turbomycin A and B, have been reported by using metagenomic approach (Gillespie et al. 2002). Traditional approach, which is screening of microbes for production of novel antibiotics, is time-consuming and labor-intensive. Various metabolomic methods like use of NMR- or MS-based metabolomics are an efficient tool to find novel antibiotics in microbial cultures (Wu et al. 2015).

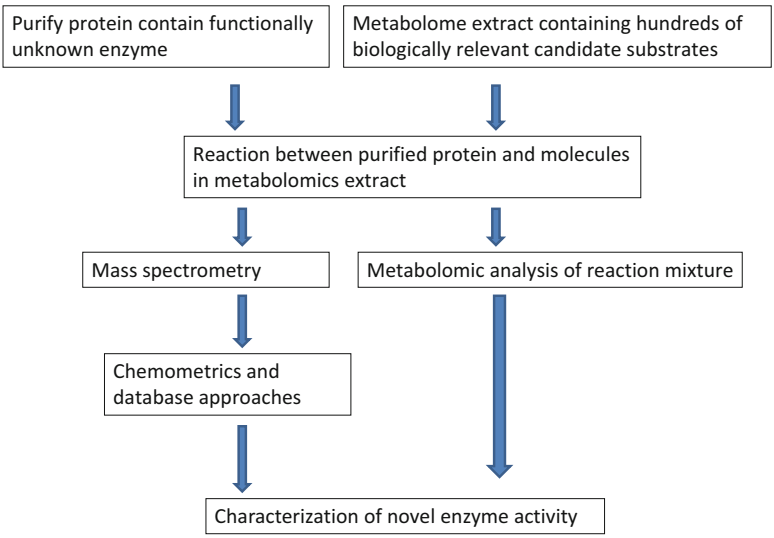


Fig. 11.3 Strategy to uncover enzymatic activities using metabolomic approach

11.8 Conclusion

Improvements in metagenomics and metabolomics methods combined with bioinformatic tools have eased the addressal of the challenges of pooling sequence data to determine the complete genome of individual community members, comparing the diversity of various environments, and assessing changes in diversity. Research into diverse microbiomes has revealed a huge amount of novelty, including genes that encode proteins which may be of significant value in maintaining the sustainability of agriculture. Interdisciplinary coordination of experts is required due to the complexity and high cost of technologies necessary for metagenomics and metabolomics. The enormous amount of data generated by computational studies is currently being made available in international archives. Future refinement of methods that enrich the genes of a particular function will help in discovery of useful molecules.

Acknowledgment PB thanks DST-SERB: SB/YS/LS-213/2013 for the financial support.

References

Alvarez-Sanchez B, Priego-Capote F, de Castro ML (2010) Metabolomics analysis II. Preparation of biological samples prior to detection. Trends Anal Chem 29(2):120–127

Béjà O, Spudich EN, Spudich JL, Leclerc M, DeLong EF (2001) Proteorhodopsin phototrophy in the ocean. Nature 411(6839):786–789

Béjà O, Koonin EV, Aravind L, Taylor LT, Seitz H, Stein JL, Bensen DC, Feldman RA, Swanson RV, DeLong EF (2002) Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. Appl Environ Microbiol 68(1):335–345

Chung EJ, Lim HK, Kim JC, Choi GJ, Park EJ, Lee MH, Chung YR, Lee SW (2008) Forest soil metagenome gene cluster involved in antifungal activity expression in *Escherichia coli*. Appl Environ Microbiol 74(3):723–730

Cottrell MT, Moore JA, Kirchman DL (1999) Chitinases from uncultured marine microorganisms. Appl Environ Microbiol 65(6):2553–2557

Courtois S, Cappellano CM, Ball M, Francou FX, Normand P, Helynck G, Martinez A, Kolvek SJ, Hopke J, Osburne MS, August PR (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. Appl Environ Microbiol 69 (1):49–55

Cullen DW, Hirsch PR (1998) Simple and rapid method for direct extraction of microbial DNA from soil for PCR. Soil Biol Biochem 30(8):983–993

Daniel R (2004) The soil metagenome–a rich resource for the discovery of novel natural products. Curr Opin Biotechnol 15(3):199–204

Defernez M, Gunning YM, Parr AJ, Shepherd LV, Davies HV, Colquhoun IJ (2004) NMR and HPLC-UV profiling of potatoes with genetic modifications to metabolic pathways. J Agric Food Chem 52(20):6075–6085

Delmont TO, Robe P, Cecillon S, Clark IM, Constancias F, Simonet P, Hirsch PR, Vogel TM (2011) Accessing the soil metagenome for studies of microbial diversity. Appl Environ Microbiol 77(4):1315–1324

Entcheva P, Liebl W, Johann A, Hartsch T, Streit WR (2001) Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. Appl Environ Microbiol 67(1):89–99

Ferrer M, Martínez-Abarca F, Golyshin PN (2005) Mining genomes and 'metagenomes' for novel catalysts. Curr Opin Biotechnol 16(6):588–593

Gidman E, Goodacre R, Emmett B, Smith AR, Gwynn-Jones D (2003) Investigating plant–plant interference by metabolic fingerprinting. Phytochemistry 63(6):705–710

Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR, Clardy J, Goodman RM, Handelsman J (2002) Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. Appl Environ Microbiol 68(9):4301–4306

Henne A, Daniel R, Schmitz RA, Gottschalk G (1999) Construction of environmental DNA libraries in *Escherichia coli* and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. Appl Environ Microbiol 65(9):3901–3907

Henne A, Schmitz RA, Bömeke M, Gottschalk G, Daniel R (2000) Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. Appl Environ Microbiol 66(7):3113–3116

Hoffmann JF, Carvalho IR, Barbieri RL, Rombaldi CV, Chaves FC (2017) Butia spp.(Arecaceae) LC-MS-based metabolomics for species and geographical origin discrimination. J Agric Food Chem 65(2):523–532

Hofreiter M, Rabeder G, Jaenicke-Després V, Withalm G, Nagel D, Paunovic M, Jambrěsić G, Pääbo S (2004) Evidence for reproductive isolation between cave bear populations. Curr Biol 14 (1):40–43

Holben WE, Jansson JK, Chelm BK, Tiedje JM (1988) DNA probe method for the detection of specific microorganisms in the soil bacterial community. Appl Environ Microbiol 54 (3):703–711

Kim UJ, Shizuya H, de Jong PJ, Birren B, Simon MI (1992) Stable propagation of cosmid sized human DNA inserts in an F factor based vector. Nucleic Acids Res 20(5):1083–1085

Kim TK, Hewavitharana AK, Shaw PN, Fuerst JA (2006) Discovery of a new source of rifamycin antibiotics in marine sponge actinobacteria by phylogenetic prediction. Appl Environ Microbiol 72(3):2118–2125

Knietsch A, Bowien S, Whited G, Gottschalk G, Daniel R (2003) Identification and characterization of coenzyme B12-dependent glycerol dehydratase-and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. Appl Environ Microbiol 69 (6):3048–3060

Lenz EM, Weeks JM, Lindon JC, Osborn D, Nicholson JK (2005) Qualitative high field ^{1} H-NMR spectroscopy for the characterization of endogenous metabolites in earthworms with biochemical biomarker potential. Metabolomics 1(2):123–136

Li ZY, Liu Y (2006) Marine sponge Craniella austrialiensis-associated bacterial diversity revelation based on 16S rDNA library and biologically active Actinomycetes screening, phylogenetic analysis. Lett Appl Microbiol 43(4):410–416

Li Y, Wexler M, Richardson DJ, Bond PL, Johnston AW (2005) Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of *Rhizobium leguminosarum* and of *Escherichia coli* reveals different classes of cloned trp genes. Environ Microbiol 7 (12):1927–1936

Lorenz P, Liebeton K, Niehaus F, Eck J (2002) Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. Curr Opin Biotechnol 13(6):572–577

MacNeil IA, Tiong CL, Minor C, August PR, Grossman TH, Loiacono KA, Lynch BA, Phillips T, Narula S, Sundaramoorthi R, Tyler A (2001) Expression and isolation of antimicrobial small molecules from soil DNA libraries. J Mol Microbiol Biotechnol 3(2):301–308

Majerník A, Gottschalk G, Daniel R (2001) Screening of environmental DNA libraries for the presence of genes conferring Na+ (Li+)/H+ antiporter activity on *Escherichia coli*: characterization of the recovered genes and the corresponding gene products. J Bacteriol 183 (22):6645–6653

Martinez A, Kolvek SJ, Yip CL, Hopke J, Brown KA, MacNeil IA, Osburne MS (2004) Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. Appl Environ Microbiol 70(4):2452–2463

Nazir A (2016) Review on metagenomics and its applications. Imp J Interdiscip Res 2(3):277–286

Noble AD, Ruaysoongnern S (2010) The nature of sustainable agriculture. In: Soil microbiology and sustainable crop production. Springer, Dordrecht, pp 1–25

Noguchi H, Park J, Takagi T (2006) MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. Nucleic Acids Res 34(19):5623–5630

Ogram A, Sayler GS, Barkay T (1987) The extraction and purification of microbial DNA from sediments. J Microbiol Methods 7(2–3):57–66

Piel J (2011) Approaches to capturing and designing biologically active small molecules produced by uncultured microbes. Annu Rev Microbiol 65:431–453

Prosser GA, Larrouy-Maumus G, de Carvalho LP (2014) Metabolomic strategies for the identification of new enzyme functions and metabolic pathways. EMBO Rep 15(6):657–669

Pinu FR, Villas-Boas SG (2017) Extracellular microbial metabolomics: the state of the art. Metabolites 7(3):43

Quaiser A, Ochsenreiter T, Lanz C, Schuster SC, Treusch AH, Eck J, Schleper C (2003) Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. Mol Microbiol 50(2):563–575

Ram RJ, VerBerkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC, Shah M, Hettich RL, Banfield JF (2005) Community proteomics of a natural microbial biofilm. Science 308 (5730):1915–1920

Riesenfeld CS, Goodman RM, Handelsman J (2004) Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. Environ Microbiol 6(9):981–989

Robe P, Nalin R, Capellano C, Vogel TM, Simonet P (2003) Extraction of DNA from soil. Eur J Soil Biol 39(4):183–190

Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL (2000) Cloning the soil metagenome: a strategy for

accessing the genetic and functional diversity of uncultured microorganisms. Appl Environ Microbiol 66(6):2541–2547
Stein JL, Marsh TL, Wu KY, Shizuya H, DeLong EF (1996) Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. J Bacteriol 178(3):591–599
Saito N, Robert M, Kitamura S, Baran R, Soga T, Mori H, Nishioka T, Tomita M (2006) Metabolomics approach for enzyme discovery. Seikagaku J Jpn Biochem Soc 83 (11):1039–1043
Schloss PD, Handelsman J (2003) Biotechnological prospects from metagenomics. Curr Opin Biotechnol 14(3):303–310
Schloss PD, Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Environ Microbiol 71 (3):1501–1506
Schmeisser C, Steele H, Streit WR (2007) Metagenomics, biotechnology with non-culturable microbes. Appl Microbiol Biotechnol 75(5):955–962
Schmid A, Dordick JS, Hauer B, Kiener A, Wubbolts M, Witholt B (2001) Industrial biocatalysis today and tomorrow. Nature 409(6817):258–268
Schripsema J (2010) Application of NMR in plant metabolomics: techniques, problems and prospects. Phytochem Anal 21(1):14–21
Shuman JL, Cortes DF, Armenta JM, Pokrzywa RM, Mendes P, Shulaev V (2011) Plant metabolomics by GC-MS and differential analysis. Plant Reverse Genet Methods Protocol:229–246
Soo EC, Aubry AJ, Logan SM, Guerry P, Kelly JF, Young NM, Thibault P (2004) Selective detection and identification of sugar nucleotides by CE− electrospray-MS and its application to bacterial metabolomics. Anal Chem 76(3):619–626
Spencer DH, Kas A, Smith EE, Raymond CK, Sims EH, Hastings M, Burns JL, Kaul R, Olson MV (2003) Whole-genome sequence variation among multiple isolates of *Pseudomonas aeruginosa*. J Bacteriol 185(4):1316–1325
Strous M, Pelletier E, Mangenot S, Rattei T, Lehner A, Taylor MW, Horn M, Daims H, Bartol-Mavel D, Wincker P, Barbe V (2006) Deciphering the evolution and metabolism of an anammox bacterium from a community genome. Nature 440(7085):790–794
Suyal DC, Yadav A, Shouche Y, Goel R (2015) Diversified diazotrophs associated with the rhizosphere of Western Indian Himalayan native red kidney beans (*Phaseolus vulgaris* L.). 3 Biotech 5(4):433–441
Teeling H, Glöckner FO (2012) Current opportunities and challenges in microbial metagenome analysis—a bioinformatic perspective. Brief Bioinform 13(6):728–742
Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428(6978):37–43
Uchiyama T, Abe T, Ikemura T, Watanabe K (2005) Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. Nat Biotechnol 23 (1):88–93
Valverde A, Aguilera A, Ferrer C, Camacho F, Cammarano A (2010) Analysis of forchlorfenuron in vegetables by LC/TOF-MS after extraction with the buffered QuEChERS method. J Agric Food Chem 58(5):2818–2823
Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304(5667):66–74
Villas-Bôas SG, Bruheim P (2007) Cold glycerol–saline: the promising quenching solution for accurate intracellular metabolite analysis of microbial cells. Anal Biochem 370(1):87–97
Voget S, Leggewie C, Uesbeck A, Raasch C, Jaeger KE, Streit WR (2003) Prospecting for novel biocatalysts in a soil metagenome. Appl Environ Microbiol 69(10):6235–6242

Wang GY, Graziani E, Waters B, Pan W, Li X, McDermott J, Meurer G, Saxena G, Andersen RJ, Davies J (2000) Novel natural products from soil DNA libraries in a streptomycete host. Org Lett 2(16):2401–2404

Wang Y, Xu L, Shen H, Wang J, Liu W, Zhu X, Wang R, Sun X, Liu L (2015) Metabolomic analysis with GC-MS to reveal potential metabolites and biological pathways involved in Pb & Cd stress response of radish roots. Sci Rep 5:18296

Williamson LL, Borlee BR, Schloss PD, Guan C, Allen HK, Handelsman J (2005) Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. Appl Environ Microbiol 71(10):6335–6344

Wilson ID, Plumb R, Granger J, Major H, Williams R, Lenz EM (2005) HPLC-MS-based methods for the study of metabonomics. J Chromatogr B 817(1):67–76

Walter J, Mangold M, Tannock GW (2005) Construction, analysis, and β-glucanase screening of a bacterial artificial chromosome library from the large-bowel microbiota of mice. Appl Environ Microbiol 71(5):2347–2354

Wu C, Kim HK, van Wezel GP, Choi YH (2015) Metabolomics in the natural products field–a gateway to novel antibiotics. Drug Discov Today Technol 13:11–17

Xing MN, Zhang XZ, Huang H (2012) Application of metagenomic techniques in mining enzymes from microbial communities for biofuel synthesis. Biotechnol Adv 30(4):920–929

Chapter 12
In Silico Plant Growth Metabolites

Shradha Sharma and Pushpa Lohani

Abstract The exponential development in civilization of human race demands for more reliable resources and technology, to sustain and secure the future of the coming generations. With the growing pace of evolution, we need highly efficient technique to scrutinize the data and deduct the possible outcomes. System biology has opened up a totally new era of analysing these complex records of living beings. The in silico on literal terms means "performed on computer or via computer simulation". The computer models of crops like rice (*Oryza sativa*), maize (*Zea mays*), soybean (*Glycine max*) and cassava (*Manihot esculenta*) have been made to study and design crops with higher yield and efficiency. This in silico analysis is a combat against the time taking crop breeding technique and error-prone genetic engineering where the result is unpredictable due to intricate interaction between genotype and management. The digital representation of plants in silico (Psi) will examine the possible phenotype of the crop from genotype and environmental interactions. In this chapter, we give a brief on secondary metabolites produced by plants and their potential applications along with the databases which have been made to easily retrieve the required data about them for scientific and academic purposes.

Keywords Secondary metabolites · Phytochemicals · Database · Simulation · Computer models

12.1 What Are Secondary Metabolites?

Organic compounds synthesized by plants are mainly classified as primary and secondary metabolites. The primary metabolites are required for fundamental procedures like respiration, photosynthesis, growth and development (Gandhi et al. 2015). These

S. Sharma · P. Lohani (✉)
Department of Molecular Biology and Genetic Engineering, College of Basic Science and Humanities, GB Pant University of Agriculture & Technology, Pantnagar, India

D. K. Choudhary et al. (eds.), *In Silico Approach for Sustainable Agriculture*,
https://doi.org/10.1007/978-981-13-0347-0_12

primary metabolites when modified by methylation, hydroxylation or glycosylation give rise to secondary metabolites (Xue and He 2015). The secondary metabolites are not the intermediates of the chemical reactions but are rather collected in plants to perform other functions. These secondary metabolites are phytochemicals which are particular to a particular species or genera and help a plant to adapt to its environment and are a source of pharmaceutical as well (Xue and He 2015; Kennedy and Wightman 2011). Importance of secondary metabolites for plants are listed below.

I. As plants are stationary living beings, they have to develop some immunity mechanism to protect themselves. These secondary metabolites are protection against microorganisms such as bacteria, fungi and viruses (Kennedy and Wightman 2011).
II. Apart from microorganisms, more complex animals like herbivores are also required to keep away. So, some of these phytochemicals are toxic for herbivores as they directly attack the central and peripheral nervous system of herbivores (Gandhi et al. 2015; Kennedy and Wightman 2011).
III. These secondary metabolites also protect a plant from abiotic stresses such as infections, UV radiation, wounding and ozone (Gandhi et al. 2015; Xue and He 2015; Kennedy and Wightman 2011).
IV. Phenylpropanoids (i.e. verbascoside, forsythoside, arenareoside, ballotetraside and caffeoyl malic acid), which are the largest group of secondary metabolites, act as powerful antioxidants against reactive oxygen and nitrogen species (ROS/RNS). Excessive ROS/RNS species may cause oxidative stress, cell death and loss in cell function. Hence, neutralizing these reactive species is important to maintain normal cellular functions (Xue and He 2015).
V. Some secondary metabolites are allelopathic defenders of competitor plants growing in the vicinity of the concerned plant. Thus, secondary metabolites also maintain an inter-plant relationship (Kennedy and Wightman 2011).
VI. The secondary metabolites are responsible for the scent and colour, which act as highly attractive features for insects. These insects help in pollination (Gandhi et al. 2015; Kennedy and Wightman 2011).

So, it can be seen from above that secondary metabolites do not play any role in plant's fundamental metabolic pathways but are rather necessary for their survival. Also these metabolites have been found to have uses for humans too. They can be used as antipyretics, hallucinogens, cholesterol-lowering drugs, herbicides, food additives, fragrances and also plastics synthesis precursors (Breitling et al. 2013). For example, *Perilla*, a medicinal plant, secretes perillaldehyde which has anti-allergic, anti-inflammatory and antidepressant effects (Lu et al. 2017). Due to the huge potential they behold, a separate section of synthetic biology, metabolomics has been established for their discovery, characterization, study and production (Breitling et al. 2013).

12.2 In Silico Modelling of Plants

Projects like Virtual Physiological Human (VPH) and Virtual Physiological Rat (VPR) have been developed around a central framework for medicinal purposes via integrative multi-scale modelling. In silico modelling of organism is a solution to many problems due to its predictive capabilities (Marshall-Colon et al. 2017). Plants are one of the most highly complex multicellular eukaryotes, which respond to their environmental stimulus like high radiation, drought, salinity and extreme temperature (Eckardt 2012). These factors limit their potential and lead to yield losses. Predicting the phenotype of a crop can be tricky due to interaction of different genotype with environment and management. As these highly complex eukaryotes are utilized for various utilities like food, medicine, cosmetics and wood by humans, maintaining a good crop yield is the demand of exponentially growing population. The unpredictable losses due to environmental factors can be avoided with the help of virtual models (Marshall-Colon et al. 2017). In silico modelling of crops will help the researchers to select potential phenotypes for different environmental stimulus based on evolutionary algorithms, Pareto efficiency and other algorithms to obtain the desired output. Also, interaction of traits with other countenance traits may lead to large number of possible phenotypic combinations. For example, if there are ten individual traits present in dominant and recessive states, then there exist two^{11} phenotypes due to combination of these traits. Evaluating so many phenotypes will be a very tedious job for scientists. In silico modelling will help them to develop robust ideotypes by narrowing the combinations based on the desired environmental stimulus (Zhu et al. 2016).

Models of C3 and C4 photosynthetic process, stomatal action, 3D plant canopies, phloem and xylem flow, respiration, growth and development, root structure and functional dynamics and flowering have been developed for the stimulation of biological process and observe the phenotype of crops in response to environmental factors. In silico modelling of plants faces some challenges. Some of them are:

I. Unlike humans, plants have many species of food and bioenergy crops which vary greatly in their phenology, plant architecture, primary and secondary metabolism and require different stimulation models for accurate results.
II. Different labs produce their own version of models which are written in different scripting languages, hence becoming redundant. Apart from this, many models are not accessible to public and also lack good quality data (Marshall-Colon et al. 2017).

12.3 Metabolite Database

Due to increase in genomic, epigenomic, metabolomic, phenomic, transcriptomic and proteomic data, there is a need of bioinformatic tools to manage it. The biological database should be an open-source software with Internet connectivity to favour easy sharing. The framework of the model should be hierarchical with

interoperability to accept data from researchers. The reproducibility of data will help the researches to submit their studies with other communities (Marshall-Colon et al. 2017; Zhu et al. 2016). Some of the useful models and data repositories are:

I. *MetaCrop*: It is a hand-curated database, based on DBMS oracle. This database has information about metabolic pathways, including their transport processes, location information and reaction kinetics. Information about 40 major metabolic pathways of both dicotyledons and monocotyledons along with reactions, the information about enzymes (EC and CAS number), metabolites (CAS number, chemical formula and molecular weight), stoichiometry and detailed location (species, tissue, organ, developmental stage and compartment) is also incorporated. There is also some manually extracted data from online databases like KEGG PATHWAY, EGENES, AraCyc, MetaCyc, RiceCyc, BRENDA and ARAMEMNON (Grafahrend-Belau et al. 2008) (Fig. 12.1).

II. *KNApSAck*: Secondary metabolites is a diverse family of compounds having 30,000 terpenoids, 9000 flavonoids, 1600 isoflavonoids and 12,000 alkaloids. These secondary metabolites are used in medicine to enhance the health quality of humans. They play a role in nutrigenomics, nutrigenetics, foodomics and medical genomics. Understanding and maintaining the database will help us understand the relationship between these metabolites so that we can further utilize them. In this database, metabolites are classified on the basis of their taxonomy and are searched on the basis of seven aspects. These are C_ID (KNApSAck compound ID), activity category, metabolite name, individual biological activity, molecular formula, target species and CAS ID. This database has metabolite data of not only plants but of microbes too (Nakamura et al. 2014; Weber and Kim 2016; Johnson and Lange 2015). Below is a list of 140 activity categories in KNApSAck (Nakamura et al. 2014) (Fig. 12.2).

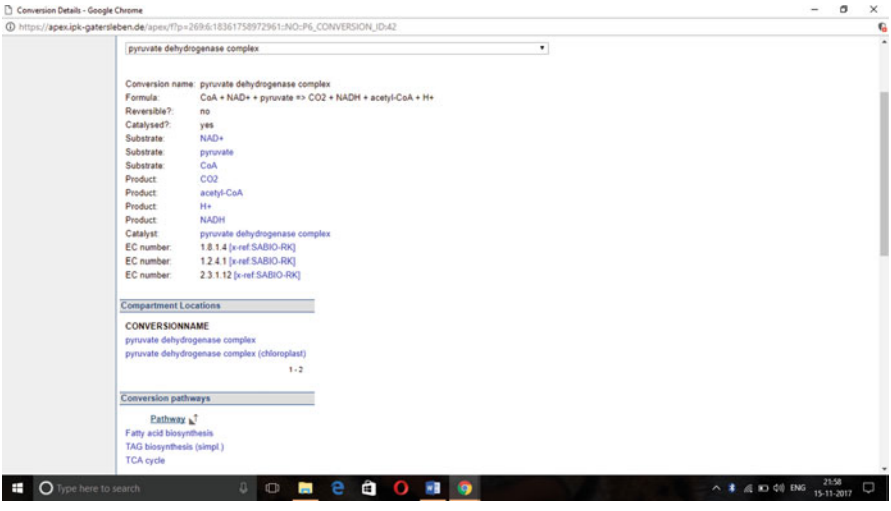


Fig. 12.1 Pyruvate dehydrogenase complex information on MetaCrop

General description	Activity category	No. of records
Plant growth regulator (946)	Enhance germination (E01)	26
	Enhance stem growth (E02)	138
	Enhance root growth (E03)	46
	Enhance leaf growth (E04)	48
	Enhance flowering (E05)	42
	Enhance fruiting (E06)	44
	Enhance plant growth (E07)	124
	Inhibit seed germination (E08)	27
	Inhibit stem growth (E09)	51
	Inhibit root growth (E10)	32
	Inhibit leaf growth (E11)	29
	Inhibit flowering (E12)	12
	Inhibit fruiting (E13)	3
	Inhibit plant growth (E14)	91
	Allelopathic (E15)	70
	Phytoalexin (E16)	163
Attractant/ repellent (236)	Feeding attractant (E17)	33
	Feeding deterrent (E18)	81
	Pollinator attractant (E19)	44
	Oviposition attractant (E20)	26
	Oviposition deterrent (E21)	3
	Sex attractant (E22)	22
	Attractant (E23)	13
	Repellent (E24)	14
Selective toxicity (163)	Phytotoxic (E25)	45
	Herbicidal (E26)	6
	Insecticidal (E27)	78
	Acaricidal (E28)	2
	Molluscicidal (E29)	12
	Piscicidal (E30)	15
	Nematocidal (E31)	5
Antimicrobial agent (1238)	Antibacterial (E32)	692
	Antituberculosis (E33)	42
	Antileprotic (E34)	7
	Antifungal (E35)	407
	Inhibit spore germination (E36)	15
	Antimicrobial (E37)	75
Antiviral agent (106)	Antiviral (E38)	83
	Antihepatitic (E39)	4
	Anti-HIV (E40)	7
	Anti-HSV (E41)	12

(continued)

General description	Activity category	No. of records
Antiparasitic agent (217)	Anthelmintic (E42)	46
	Antiprotozoal (E43)	7
	Antiamebic (E44)	12
	Antimalarial (E45)	64
	Antileishmanial (E46)	32
	Antitrypanosomal (E47)	54
	Pediculicide (E48)	2
Nervous system agent (443)	Antipyretic (M01)	31
	Analgesic (M02)	66
	Antiarthritic (M03)	4
	Anaesthetics (M04)	21
	Sedative (M05)	75
	Antispasmodic (M06)	60
	Anticonvulsant (M07)	12
	Antidementic (M08)	15
	Antidepressant (M09)	10
	CNS stimulant (M10)	30
	Diaphoretic (M11)	6
	Emetic (M12)	5
	Antiemetic (M13)	7
	Antigout (M14)	5
	Antimigraine (M15)	4
	Antimyasthenic (M16)	8
	Antiparkinson (M17)	26
	Antipsychotic (M18)	25
	Muscle relaxant (M19)	33
Cardiovascular agent (398)	Antidiabetic (M20)	58
	Hemostatic (M21)	8
	Antithrombotic (M22)	2
	Cardiotonic (M23)	38
	Antiarrhythmic (M24)	18
	Diuretic (M25)	16
	Antihypertensive (M26)	165
	Antihyperlipidemic (M27)	29
	Antianemic (M28)	4
	Other cardiovascular agent (M29)	60
Respiratory tract agent (108)	Antitussive (M30)	29
	Expectorant (M31)	13
	Antiasthmatic (M32)	57
	Other respiratory tract agent (M33)	9

(continued)

General description	Activity category	No. of records
Digestive organ agent (168)	Antidiarrheic (M34)	12
	Carminative (M35)	5
	Stomachic (M36)	3
	Laxative (M37)	38
	Choleretic (M38)	17
	Antihepatotoxic (M39)	49
	Other digestive organ agent (M40)	44
Genitourinary agent (54)	Oxytocic (M41)	21
	Antifertility (M42)	22
	Abortifacient (M43)	5
	Other genitourinary agent (M44)	6
Anticancer agent (495)	Antioxidant (M45)	94
	Anticancer (M46)	267
	Antitumour (M47)	83
	Antineoplastic (M48)	38
	Antimutagenic (M49)	13
Anti-inflammatory agent (257)	Anti-inflammatory (M50)	149
	Anti-allergic (M51)	57
	UV shield (M52)	7
	Antidermatitic (M53)	19
	Antiedemic (M54)	25
Immunological agent (31)	Immunosuppressant (M55)	11
	Immunostimulant (M56)	11
	Immunomodulator (M57)	9
Nutrient (76)	Nucleic acid (M58)	4
	Essential amino acid (M59)	15
	Nonessential amino acid (M60)	8
	Vitamin (M61)	14
	Nutrient (M62)	25
	Tonic (M63)	10
Nontherapeutic agent (332)	Solvent (M64)	4
	Flavour (M65)	127
	Odour (M66)	67
	Pigment (M67)	98
	Emulsifying agent (M68)	2
	Antiseptic (M69)	34
Other health agent (195)	Antiulcerogenic (M70)	25
	Depilatory (M71)	6
	Antidote (M72)	17
	Hormonal (M73)	13
	Dental (M74)	5
	Other health agent (M75)	129

(continued)

General description	Activity category	No. of records
Narcotic (14)	Narcotic (M76)	14
Toxic (640)	Phototoxic (M77)	33
	Neurotoxic (M78)	47
	Pneumotoxic (M79)	27
	Hepatotoxic (M80)	41
	Cytotoxic (M81)	57
	Toxic (M82)	435
Tumorigenic (79)	Tumorigenic (M83)	24
	Mutagenic (M84)	32
	Genotoxic (M85)	4
	Teratogenic (M86)	19
Other disease-causing agent (159)	Psychotomimetic (M87)	23
	Hemolytic (M88)	52
	Allergenic (M89)	23
	Irritant (M90)	37
	Dermatitic (M91)	21
	Edematous (M92)	3

III. *MetaCyc*: It is another highly curated database metabolic pathways and enzymes, incorporating data from all domains of life. It has data of 445 pathways, 1115 enzymes of 158 arising in organisms. Scientists all over the world can improve the data. It is also linked with other databases like MetaCyc; proteins are linked with PROSITE, PRINTS, PANTHER, Pfam, SWISS PROT, UNIPROT, MODBASE, DIP, etc. MetaCyc gene is linked with NCBI-gene, STRING, ECOGENE, ECOBASE, RefSeq, MIM, ArrayExpress, etc. Pathway genome databases of many organism, including *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Oryza sativa*, *Mus musculus*, *Bos Taurus*, *Medicago truncatula*, *Populus trichocarpa*, *Dictyostelium discoideum*, *Leishmania major*, *Chlamydomonas reinhardtii* and many more, have been made with the help of MetaCyc (Caspi et al. 2013; Karp et al. 2002) (Fig. 12.3).

IV. *DoBISCUIT (database of biosynthesis clusters curated and integrated)*: This database focuses on secondary metabolites obtained from bacteria, especially actinomytes. The main content in DoBISCUIT focuses on biosynthetic clusters (cluster information page). This page has six sections, namely, genomic map, compound original source, PKS/NRPS modules, data download and references. The compound section exhibits the chemical structure, biological activities and structural attributes such as chain length and sugar attachment. Compound original section displays the bacterial strain from which the biosynthetic clusters were obtained. The PKS/NRPS section displays the domain association of each module in the enzyme. The reference section displays the links for references from where the data has been collected, and data download section

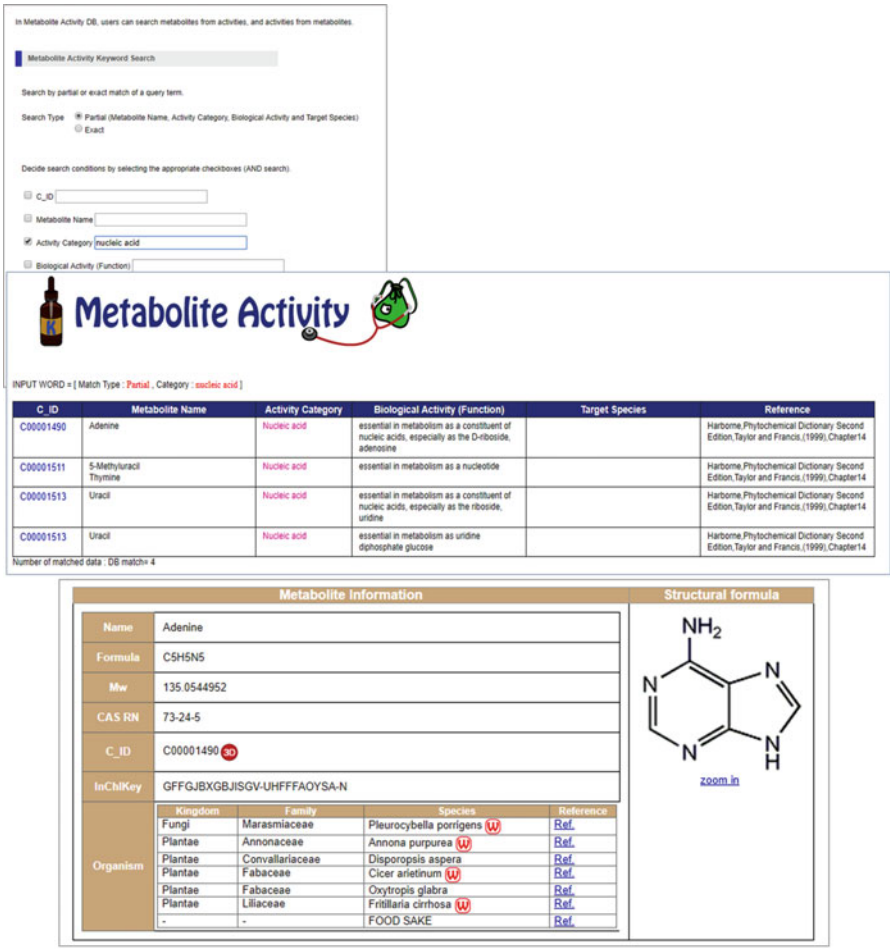


Fig. 12.2 Screenshots of webpages showing results of adenine, a molecule of nucleic acid activity category

is for downloading things such as nucleotide sequence of clusters, CDS nucleotide/amino acid sequences in multi-FASTA format and curated comments in CSV or GenBank format (Ichikawa et al. 2013) (Fig. 12.4).

V. *NeeMDB*: Neem (*Azadirachta indica*) is a medicinally important plant of Indian subcontinent. The extracts of neem from seeds, kernels and barks have antifeedant, repellent, anti-ovipositional, insect growth regulatory, fecundity and fitness-reducing properties on insects. These properties are due to a variety of secondary metabolites displayed by neem. Some important ones are

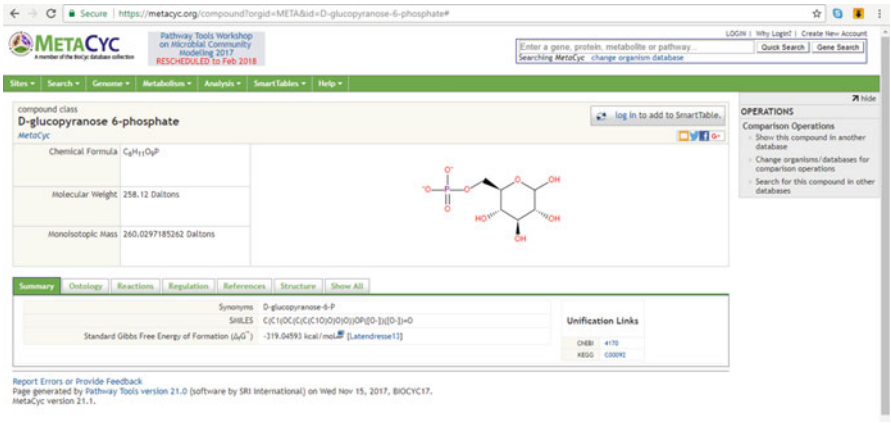


Fig. 12.3 Screenshot of webpage showing information about D-glucopyranose 6-phosphate

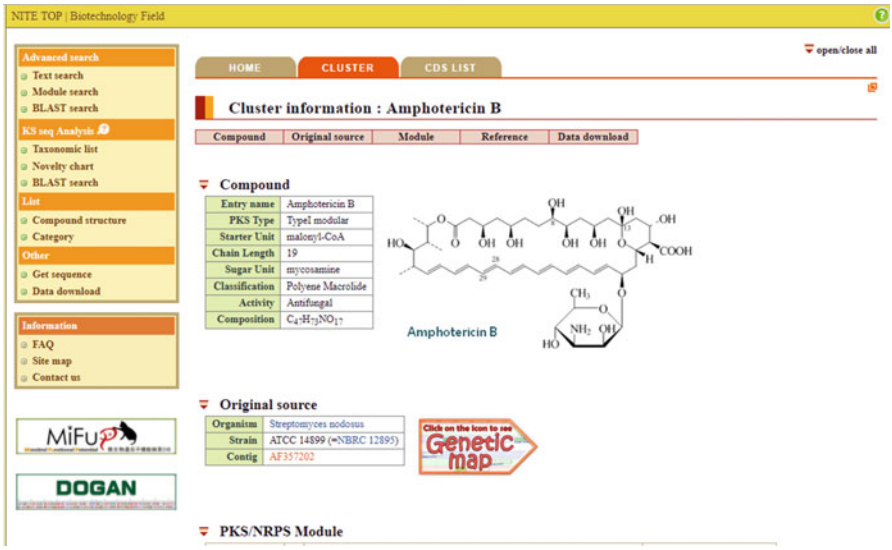


Fig. 12.4 Screenshot of webpage showing results for amphotericin B

azadirone, meliacarpin, gedunin, salanin, nimbin and vilasinin showing medicinal and pesticidal properties. NeeMDB has 250 secondary metabolites in its database. These metabolites are being searched on the basis of alphabetically, molecular weight, structure, IUPAC name or functional group. Firefox 10.0 or above, chrome 27.0 and Java 6 or above are the minimum requirements for this database (Hatti et al. 2014) (Fig. 12.5).

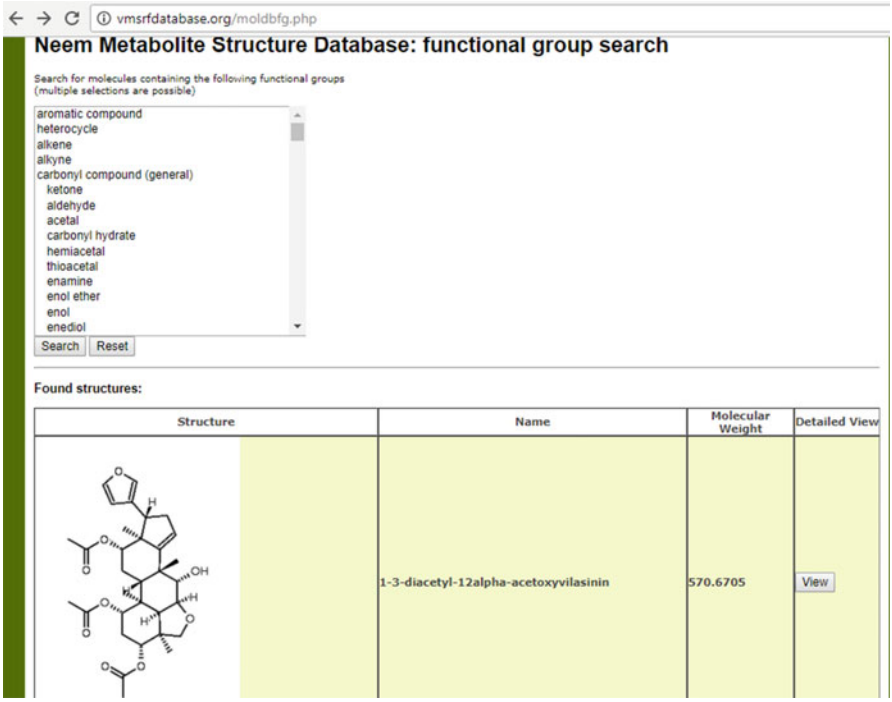


Fig. 12.5 Screenshot of webpage of NeeMDB showing results for aromatic compounds

VI. *SWMD (seaweed metabolite database)*: Marine life has huge potential for sources which are valuable for pharmaceuticals. The seaweed has approximately 3300 metabolites. Other algae like brown and red algae are fertile producers of secondary metabolites like sesquiterpenes, diterpenes, triterpenes and C15 acetogenins. An open-access database of marine metabolites will be useful for researchers around the world to study the natural products obtained. SWMD provides scientific community a quick and easy way to analyse the metabolites, helping them in drug discovery.

The compounds in the database are interpreted on the basis of molecular weight, molar refractivity, monoisotopic mass, number of rotatable bonds, number of hydrogen bond donors, calculated logP, number of hydrogen bond receptors, van der Waals surface area and polar surface area. The 3D structure for each compound is also available, which is calculated using MarvinSketch and ChemSketch. The compounds in the database can be searched on the basis of compound name, accession number, seaweed binomial name, IUPAC name, compound type, InChI or SMILES. The SWMD currently endows 517 compounds, in 25 eloquent fields (Davis and Vasanthi 2011) (Figs. 12.6 and 12.7).

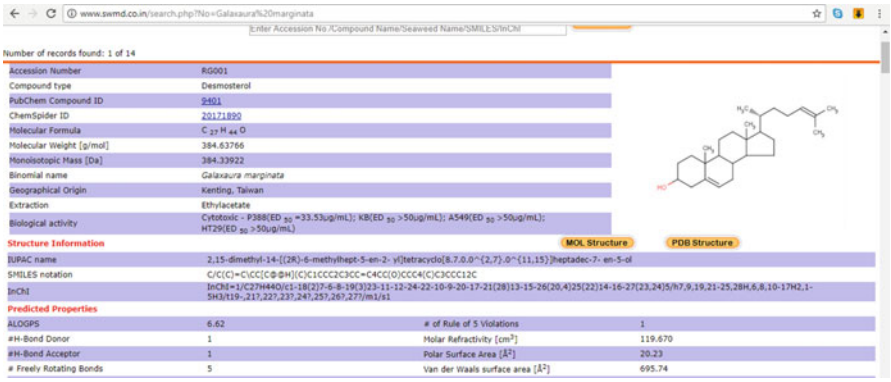


Fig. 12.6 Screenshot of webpage showing results for the entries of *Galaxaura marginata*

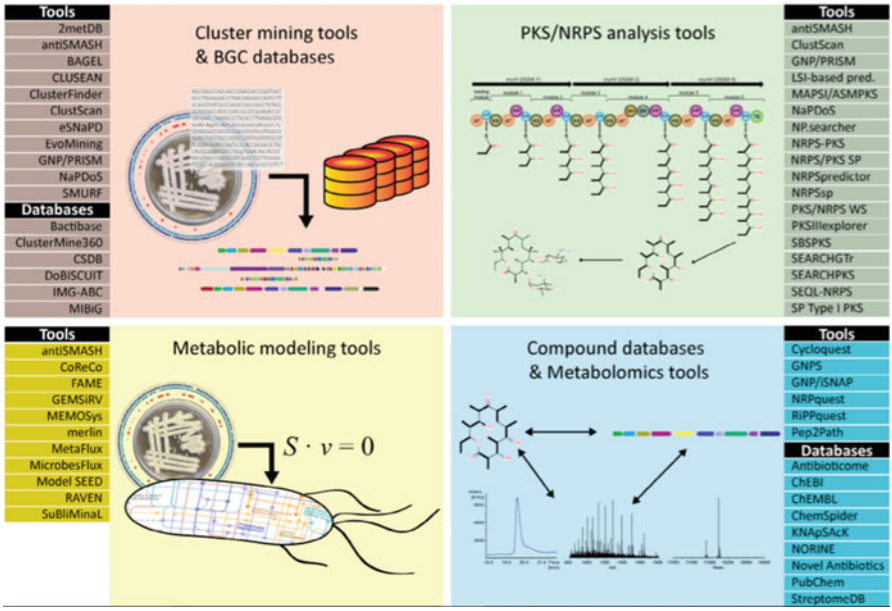


Fig. 12.7 An overview of tools and databases for secondary metabolites (Weber and Kim 2016)

12.4 Applications and Production of Secondary Metabolites

Secondary metabolites, as described above, are produced as intermediates of fundamental reactions. They are not required for plant growth but are rather important for other survival mechanisms (Gandhi et al. 2015). These secondary metabolites are also found to be useful for humans as drugs for the treatment of biological disorders.

Alternative activities of antibiotics.

Antibiotic group	Unexpected activities
β-lactams	Antitumor (inducing apoptosis)
Tetracyclines	Antiprion, antimalaria
Aminoglycosides and macrolides	Pulmonary disease, immunomodulation, antitumor, antiparastic (leishmaniasis), antimalaria
Chloropeptins	Anti-HIV
Lincosamides	Antimalaria
Isoxazolidinones	Neurotransmission
Prodigines	Antiprotozoa, antimalaria, anticancer, immunosuppression
Polyenes	Imunomodulation, antiprion, antiviral, antitumor, anti-HIV, antiparasitic (leishmaniasis)
Coumermycins	Antitumor
Glycopeptides	Anti-HIV
Ansamycins	Anticancer, antiviral
Violacein	Antiprotozoal, antitumor, antiviral, anti-ulcer
Fosfidomycin	Antimalaria

Fig. 12.8 Medicinal use of antibiotic groups (Vaishnav and Demain 2010)

They can act as antibacterial, antifungal, anticancer, immunosuppressant, cholesterol-lowering agents, herbicides, antiparasitic agents, diagnostics and tools for research. Apart from this, many metabolites are shown to have anti-HIV, anti-ageing, antihelmintic and antiprotozoal activities. Potential uses of secondary metabolites for humans are listed below (Vaishnav and Demain 2010).

(a) *Antibiotics*: Now-a -days, many antibiotics have been developed that can be used as anticancer, antimalarial, inhibition of angiogenesis and many more. A list of functions performed by different groups of antibiotics is shown below in table (Fig. 12.8).

(b) *Antitumour agents*: Metabolites class like that of anthracycline has potential antitumour effects. The main members of this class are doxorubicin, bleomycin, daunorubicin, epirubicin, valrubicin and dactinomycin. The bleomycin member also inhibits the replication of hepatitis C virus and is also active against HIV. Taxol, which is an alkaloid, is also used for cancer treatments. Bryostatins are under the trials for cancer treatments.

(c) *Immunosuppresants*: Rapamycin, tacrolimus and cyclosporine A have been used as antifungal and immunosuppressant as well. They are used in the domain of organ transplantation. Apart from this, rapamycin has also been used as anticancer agent and cardiology. Also, cyclosporine A has been used for the treatment of asthma.

(d) *Cholesterol-lowering agents*: Metabolites like compactin and lovastatin have been found to have cholesterol-lowering effects. They are effective in lowering serum LDL (low-density lipoprotein).
(e) *Other uses*: Shikonin, a type of pigment, is used in cosmetic industry. It also has anti-HIV, antitumour, anti-inflammatory and wound-healing properties.

Sophorolipids are used in cosmetic, food, pharmaceutical and cleaning industries.

Betulinic acid acts as inhibitor of topoisomerase and hence is used as antitumour agent. It is also used as antiretroviral, anti-inflammatory and antimalarial agent.

For the extraction of secondary metabolites commercially, research has been continuously going on for the improvement of better strains and techniques. Different strains of cyanobacteria have been developed for the production of different metabolites. For example, *Synechocystis* sp. PCC 6803 strain has been engineered to produce isoprene. Similarly, one more strain of above species has been developed for the production of caffeic acid (Xue and He 2015). Apart from this, the secondary metabolites have also been produced by plant tissue culture. Production through plant tissue culture reduces labour cost, increases productivity and is also free from contamination. Metabolites such as *taxol*, *morphine and codeine*, *L-DOPA*, *diosgenin*, *capsaicin*, *camptothecin* and *berberine* have been successfully produced through plant tissue cultures (Hussain et al. 2012).

References

Breitling R, Ceniceros A, Andris J, Eriko T (2013) Metabolomics for secondary metabolite research. Metabolites 3:1076–1083

Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, Holland TA, Keseler IM, Kothari A, Kub A (2013) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acid Res 42:D459–D471

Davis, Vasanthi (2011) Bioinformation 5(8):361–364

Eckardt NA (2012) In silico plant biology comes of age. Plant Cell 24:3857–3858

Gandhi SG, Mahajan V, Bedi YS (2015) Changing trends in biotechnology of secondary metabolism in medicinal and aromatic plants. Planta 214:303–317

Grafahrend-Belau E, Weise S, Koschützki D, Scholz U, Junker BJ, Schreiber F (2008) MetaCrop: a detailed database of crop plant metabolism. Nucleic Acid Res 36:D954–D958

Hatti et al (2014) Bioinformation 10(5):314–315

Hussain MS, Fareed S, Ansari S, Rahman MA, Ahmad IZ, Saeed M (2012) Current approaches towards production of secondary plant metabolites. J Pharm Bioallied Sci 4(1):10–20

Ichikawa N, Sasagawa M, Yamamoto M, Komaki H, Yoshida Y, Yamazaki S, Fujita N (2013) DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. Nucleic Acid Res 41:D408–D414

Johnson SR, Lange BM (2015) Open access metabolomics databases for natural product research: present capabilities and future potential. Front Bioeng Biotechnol 3:22

Karp PD, Riley M, Paley SM, Toole AP (2002) The MetaCyc database. Nucleic Acid Res 30 (1):59–61

Kennedy DO, Wightman (2011) Herbal extracts and phytochemicals: plant secondary metabolites and the enhancement of human brain function. Adv Nutr 2:32–50

Lu N, Bernardo EL, Tippayadarapanich C, Takagaki M, Kagawa N, Yamori W (2017) Growth and accumulation of secondary metabolites in Perilla as affected by photosynthetic photon flux density and electrical conductivity of the nutrient solution

Marshall-Colon A, Long SP, Allen DK, Allen G, Beard DA, Benes B, von Caemmerer S, Christensen AJ, Cox DJ, Hart JC, Hirst PM, Kannan K, Katz DS, Lynch JP, Millar AJ, Panneerselvam B, Price ND, Prusinkiewicz P, Raila D, Shekar RG, Shrivastava S, Shukla D, Srinivasan V, Stitt M, Turk MJ, Voit EO, Wang Y, Yin X, Zhu X-G (2017) Crops in silico: generating virtual crops using an integrative and multi-scale modeling platform. Front Plant Sci 8:786

Nakamura Y, Afendi FM, Parvin AK, Ono N, Tanaka K, Morita AH, Sato T, Sugiura T, Altaf-Ul-Amin M, Kanaya S (2014) Plant Cell Physiol 55(1):e7. (1–9)

Vaishnav P, Demain AL (2010) Unexpected application of secondary metabolites. Biotechnol Adv 29:223–229

Weber T, Kim HU (2016) The secondary metabolite bioinformatics portal: computational tools to facilitate synthetic biology of secondary metabolite production. Synth Syst Biotechnol 1:69–79

Xue Y, He Q (2015) Cyanobacteria as cell factories to produce plant secondary metabolites. Front Bioeng Biotechnol 3:57

Zhu XG, Lynch JP, LeBauer DS, Millar AJ, Stitt M, Long SP (2016) Plants in silico: why, why now and what?—An integrative platform for plant systems biology research. Plant Cell Environ 39:1049–1057

Chapter 13
Pharmacokinetic and Molecular Docking Studies of Plant-Derived Natural Compounds to Exploring Potential Anti-Alzheimer Activity

Aftab Alam, Naaila Tamkeen, Nikhat Imam, Anam Farooqui, Mohd Murshad Ahmed, Safia Tazyeen, Shahnawaz Ali, Md Zubbair Malik, and Romana Ishrat

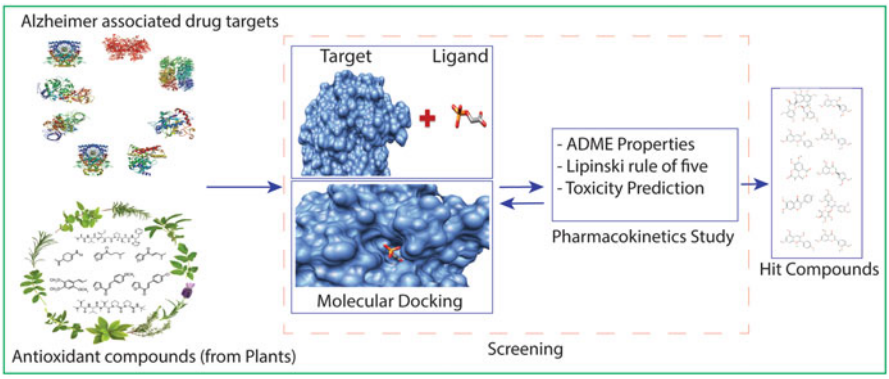
Abstract Alzheimer disease (AD) is the leading cause of dementia and accounts for 60–80% cases. Two main factors called β-amyloid (Aβ) plaques and tangles are prime suspects in damaging and killing nerve cells. However, oxidative stress, the process which produces free radicals in cells, is believed to promote its progression to the extent that it may responsible for the cognitive and functional decline observed in AD. As of today there are few FDA-approved drugs in the market for treatment, but their cholinergic adverse effect, potentially distressing toxicity and limited targets in AD pathology, limits their use. Therefore, it is crucial to find an effective compounds to combat AD. We choose 45 plant-derived natural compounds that have antioxidant properties to slow down disease progression by quenching free redicals or promoting endogenous antioxidant capacity. However, we performed molecular docking studies to investigate the binding interactions between natural compounds and 13 various anti-Alzheimer drug targets. Three known cholinesterase inhibitors (donepezil, galantamine and rivastigmine) were taken as reference drugs over natural compounds for comparison and drug-likeness studies. Few of these compounds showed good inhibitory activity besides antioxidant activity. Most of these compounds followed pharmacokinetic properties that make them potentially promising drug candidates for the treatment of Alzheimer disease.

A. Alam · N. Tamkeen · A. Farooqui · M. M. Ahmed · S. Tazyeen · S. Ali · M. Z. Malik
R. Ishrat (✉)
Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi, India
e-mail: rishrat@jmi.ac.in

N. Imam
Institute of Computer Science and Information Technology, Magadh University, Bodhgaya, India

D. K. Choudhary et al. (eds.), *In Silico Approach for Sustainable Agriculture*,
https://doi.org/10.1007/978-981-13-0347-0_13

Graphical Abstract



Pharmacokinetic and molecular docking studies of 45 natural antioxidant compounds with most known Alzheimer-associated targets

Keywords AD · Cholinesterase inhibitors · RO5 · ADMET · Docking · Pharmacology · Pharmacokinetics

13.1 Introduction

Alzheimer disease is one of the most common neurodegenerative disorders that normally cause dementia and affect the middle- to old-aged persons, around one in four individuals over the age of 85 (Feng and Wang 2012). Alzheimer is a progressive disease, where dementia steadily gets worse over the years. The World Alzheimer Report 2015 led by King's College London found that there are currently around 46.8 million people living with dementia around the world, with numbers projected to nearly double every 20 years, increasing to 74.7 million by 2030 and 131.5 million by 2050 (Fig. 13.1). At the country level, ten countries are home to over a million people with dementia in 2015: China (9.5 million), the USA (4.2 million), India (4.1 million), Japan (3.1 million), Brazil (1.6 million), Germany (1.6 million), Russia (1.3 million), Italy (1.2 million), Indonesia (1.2 million) and France (1.2 million). Thus, this condition will bring gigantic financial and personal burdens to current and future generations. In order to deal with this problem, effective therapeutic and preventive interventions should be developed urgently.

There are no such drugs/treatments available that can cure AD or any other common type of dementia completely. However, medications have been developed for Alzheimer disease that can temporarily attenuate the symptoms or delay its progression. The US Food and Drug Administration (FDA) has approved two medications: **c**holinesterase inhibitors and memantine. Over the past decade, much of the research on Alzheimer disease (AD) has focused on oxidative stress mechanisms and its importance in disease pathogenesis. The net effect of oxygen radicals is

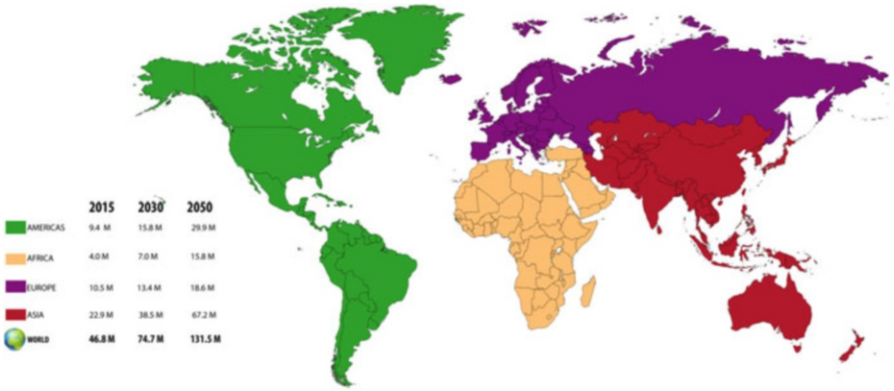


Fig. 13.1 People living with dementia around the world: 9.4 million people of America, 4.0 million of African region, 10.5 *million* of Europian countries and 22.9 *million* people of Asia are living with dementia in 2015. This no. will reach 15.8 and 29.9 *million* in Americas, 7.0 and 15.8 million in Africa, 13.4 and 18.6 *million* in Europe and 38.5 and 67.2 *million* in Asia in 2030 and 2050, respectively. (World Alzheimer report)

damaging, such damage present in AD includes advanced glycation end products (Smith et al. 1994), nitration (Smith et al. 1997), lipid peroxidation adduction products (Montine et al. 1996; Sayre et al. 1997) as well as carbonyl-modified neurofilament protein and free carbonyls (Smith et al. 1991, 1995). Significantly, this damage involves all neurons at risk to death in AD, not just those containing neurofibrillary tangles.

Nature has gifted us lots of natural remedies in the form of fruits, leaves, bark, vegetables, nuts, etc. The various ranges of bioactive nutrients present in these natural products play a vital role in prevention and cure of various neurodegenerative diseases, such as AD, Parkinson's disease and other neuronal dysfunctions. Previous studies suggested that the naturally occurring phytochemicals, such as polyphenolic antioxidants found in fruits, vegetables, herbs and nuts, may potentially hinder neurodegeneration and improve memory and cognitive functions.

In our study, we chose 45 natural compounds (antioxidant properties) of various plants from different databases (Table 13.1) for molecular docking and pharmacokinetic studies. Recent studies have demonstrated that the natural compounds and their derivatives possess wide range of biological activities like antitubercular, antifungal, antibacterial antimalarial, anti-inflammatory and antioxidant activities (Manivannan et al. 2015; Sharma et al. 2016; Choi et al. 2012). In our study, all the 45 compounds have antioxidant properties, and it has been shown that treatment with these compounds certainly contribute to their neuroprotective effects, and it is a potential approach for slowing disease progression. Therefore, we further screened our compounds against Alzheimer, which is caused by oxidative stress, and it is one of the main factors in progression of Alzheimer (Grundman et al. 2002). Natural compounds that have antioxidant properties exhibit their antioxidant effect by quenching free radicals or promoting endogenous antioxidant capacity, and some

Table 13.1 List of studied compounds and sources

S. no.	Compound	Molecular formula	Plants and food sources
1.	Vitamin C	$C_6H_8O_6$	Kakadu plum, camu camu, acerola, citrus fruits (such as oranges, sweet lime, etc.), green peppers, broccoli, green leafy vegetables, black currants, strawberries, blueberries, sea buckthorn, raw cabbage and tomatoes
2.	Vitamin E	$C_{29}H_{50}O_2$	Wheat germ, sea buckthorn, nuts, seeds, whole grains, green leafy vegetables, kiwifruit, vegetable oil and fish-liver oil
3.	Coenzyme Q10	$C_{59}H_{90}O_4$	Sardine, mackerel,beef, pork, chicken heart, chicken liver, rapeseed oil, soybean oil, sesame oil, peanuts, parsley, perilla, broccoli, grapes, cauliflower, avocados, etc.
4.	Iodide	I^-	Sea vegetables, cranberries, yoghurt, navy beans, strawberries, raw cheese, potatoes, etc.
5.	Melatonin	$C_{13}H_{16}N_2O_2$	Cherries, bananas, grapes, rice, cereals, herbs, plums, olive oil, wine, beer, pineapple, oranges, etc.
6.	Alpha-carotene	$C_{40}H_{56}$	Carrots, sweet potatoes, pumpkin, winter squash, broccoli, green beans, green peas, spinach, turnip greens, collards, leaf lettuce, avocado, parsley, etc.
7.	Astaxanthin	$C_{40}H_{52}O_4$	Microalgae, yeast, salmon, trout, krill, shrimp, crayfish, crustaceans, etc.
8.	Beta-carotene	$C_{40}H_{56}$	Butternut squash, carrots, orange bell peppers, pumpkins, kale, peaches, apricots, mango, turnip greens, broccoli, spinach and sweet potatoes
9.	Canthaxanthin	$C_{40}H_{52}O_2$	Pacific salmon, green algae, crustacean and fishes like carp, golden mullet, sea bream and rush wrasse
10.	Lutein	$C_{40}H_{56}O_2$	Spinach, kale, Swiss chard, collard greens, beet and mustard greens, endive, red pepper and okra
11.	Lycopene	$C_{40}H_{56}$	Cooked red tomato products like canned tomatoes, tomato sauce, tomato juice and garden cocktails, guava, red carrots, watermelons, gac, papayas, etc.
12.	Zeaxanthin	$C_{40}H_{56}O_2$	Paprika, corn, saffron, wolfberries, spirulina, dark green leafy vegetables, such as kale, spinach, turnip greens, collard greens, romaine lettuce, watercress, Swiss chard and mustard greens
13.	Apigenin	$C_{15}H_{10}O_5$	Parsley, celery, celeriac, chamomile tea, flowers of chamomile plants, navy bean, etc.
14.	Luteolin	$C_{15}H_{10}O_6$	*Salvia tomentosa*, celery, broccoli, green pepper, parsley, thyme, dandelion, perilla, chamomile tea, carrots, olive oil, peppermint, rosemary, navel oranges, oregano, seeds of the palm *Aiphanes aculeata*, etc.
15.	Tangeritin	$C_{20}H_{20}O_7$	Tangerine and other citrus peels
16.	Isorhamnetin	$C_{16}H_{12}O_7$	Pungent yellow or red onions, Mexican terragon (*Tagetes lucida*), etc.

(continued)

Table 13.1 (continued)

S. no.	Compound	Molecular formula	Plants and food sources
17.	Kaempferol	$C_{15}H_{10}O_6$	Apples, grapes, tomatoes, green tea, potatoes, onions, broccoli, Brussels sprouts, squash, cucumbers, lettuce, green beans, peaches, blackberries, raspberries and spinach. Plants that are known to contain kaempferol include *Aloe vera*, *Coccinia grandis*, *Cuscuta chinensis*, *Euphorbia pekinensis*, *Glycine max*, *Hypericum perforatum*, *Moringa oleifera*, *Rosmarinus officinalis*, *Sambucus nigra* and *Toona sinensis*
18.	Myricetin	$C_{15}H_{10}O_8$	Vegetables, fruits, nuts, berries, tea and is also found in red wine
19.	Quercetin	$C_{15}H_{10}O_7$	Capers, lovage, dock-like sorrel, radish leaves, carob fibre, dill, cilantro, Hungarian wax peppers, funnel leaves, onion (red), radicchio, watercress, buckwheat, kale, etc.
20.	Eriodictyol	$C_{15}H_{12}O_6$	In twigs of *Millettia duchesnei*, in *Eupatorium arnottianum* and its glycosides (eriocitrin) in lemons and rose hips (*Rosa canina*)
21.	Hesperetin	$C_{16}H_{14}O_6$	Lemons and sweet oranges
22.	Naringenin	$C_{15}H_{12}O_5$	Grapefruit, oranges, tomatoes (skin) and in water mint
23.	Catechin	$C_{15}H_{14}O_6$	*Uncaria rhynchophylla*, pome fruits, cocoa, prune juice, broad bean pod, acai oil, argan oil, peaches, green tea, vinegar, barley grain, etc.
24.	Gallocatechin	$C_{15}H_{14}O_7$	Green tea, bananas, persimmons and pomegranates
25.	Epicatechin		*Uncaria rhynchophylla*, cacao beans, green tea, etc.
26.	Epigallocatechin		St John's wort, etc.
27.	Theaflavin	$C_{29}H_{24}O_{12}$	Black tea, etc.
28.	Daidzein	$C_{15}H_{10}O_4$	Kwao Krua (*Pueraria mirifica*), Kudzu (*Pueraria lobata*), *Maackia amurensis* cell cultures, soybeans, lupin, fava beans, psoralea, *Flemingia vestita*, *F. macrophylla* and coffee
29.	Ganistein	$C_{15}H_{10}O_5$	Lupin, fava beans, soybeans, kudzu, *Flemingia vestita*, *F. macrophylla*, coffee, *Maackia amurensis* cell cultures, etc.
30.	Glycitein	$C_{16}H_{12}O_5$	Soy food products
31.	Resveratrol	$C_{14}H_{12}O_3$	Skin of grapes, blueberries, raspberries, mulberries, lingonberry, senna and concentrated in red wine
32.	Pterostilbene	$C_{16}H_{16}O_3$	Almonds, various *Vaccinium* berries, grape leaves, vines and blueberries
33.	Cyanidin	$C_{15}H_{11}O_6^+$	Grapes, bilberry, blackberry, blueberry, cherry, cranberry, elderberry, hawthorn, loganberry, açai berry, raspberry, apples, plums, red cabbage and red onion

(continued)

Table 13.1 (continued)

S. no.	Compound	Molecular formula	Plants and food sources
34.	Delphinidin	$C_{15}H_{11}O_7^+$	Cranberries, concord grapes, pomegranates, bil berries, etc.
35.	Malvidin	$C_{17}H_{15}O_7$+	*Primula* plants, blue pimpernel (*Anagallis monelli*), red wine, *Vitis vinifera,* chokeberries (*Aronia* sp), saskatoon berries (*Amelanchier alnifolia*), etc.
36.	Pelargonidin	$C_{15}H_{11}O_5^+$	Red geraniums (Geraniaceae), *Philodendron* (Araceae), flowers of blue pimpernel (*Anagallis monelli*, Myrsinaceae), raspberries, strawberries, blueberries, blackberries, cranberries, saskatoon berries, chokeberries, plums, pomegranates, kidney beans, etc.
37.	Peonidin	$C_{16}H_{13}O_6^+$	Raw cranberries, blueberries, plums, grapes, cherries, purple fleshed sweet potatoes, raw black rice and black bananas
38.	Petunidin	$C_{16}H_{13}O_7^+$ (Cl^-)	Chokeberries (*Aronia sp*), Saskatoon berries (*Amelanchier alnifolia*) or different species of grape (for instance, *Vitis vinifera*, or muscadine, *Vitis rotundifolia*), indigo rose tomatoes
39.	Cichoric acid	$C_{22}H_{18}O_{12}$	*Cichorium intybus* (chicory), *E. purpurea*, dandelion leaves, basil, lemon balm and aquatic plants, including algae and seagrasses
40.	Chlorogenic acid	$C_{16}H_{18}O_9$	Leaves of *Hibiscus sabdariffa*, potatoes, flesh of egg plants,peach, prunes, green coffee bean extract and green tea
41.	Cinnamic acid	$C_9H_8O_2$	Oil of cinnamon, balsams such as storax, shea butter, seeds of plants such as brown rice, whole wheat, oats, coffee, apple, artichoke, peanut, orange and pineapple
42.	Ellagic acid	$C_{14}H_6O_8$	North American white oak (*Quercus alba*, European red oak (*Quercus robur*). *Myriophyllum spicatum*, medicinal mushroom *Phellinus linteus*, walnuts, pecans, cranberries, raspberries, strawberries and grapes, as well as distilled beverages, peach and other plant foods
43.	Gallic acid	$C_7H_6O_5$	Parasitic plant *Cynomorium coccineum*, the aquatic plant *Myriophyllum spicatum*, the blue-green alga *Microcystis aeruginosa*. oak species, *Caesalpinia mimosoides,* and in the stem bark of *Boswellia dalzielii*, fruits (including strawberries, grapes, bananas), as well as teas, cloves and vinegars
44.	Rosmarinic acid	$C_{18}H_{16}O_8$	*Ocimum basilicum* (basil), *Ocimum tenuiflorum* (holy basil), *Melissa officinalis* (lemon balm), *Rosmarinus officinalis* (rosemary), *Origanum majorana* (marjo marjoram), *Salvia officinalis* (sage), thyme and pep permint, (*Prunella vulgaris*) *Heliotropium foertherianum Maranta* (*Maranta leuconeura*, *Maranta depressa*) and *Thalia* (*Thalia geniculata*)

(continued)

Table 13.1 (continued)

S. no.	Compound	Molecular formula	Plants and food sources
45.	Salicylic acid	$C_7H_6O_3$	Blackberries, blueberries, cantaloupes, dates, grapes, kiwi fruits, guavas, apricots, green pepper, olives, tomatoes, radish, chicory and mushrooms. Some herbs and spices contain high amounts of the legumes, seeds, nuts and cereals; only almonds, water chestnuts and peanuts have significant amounts

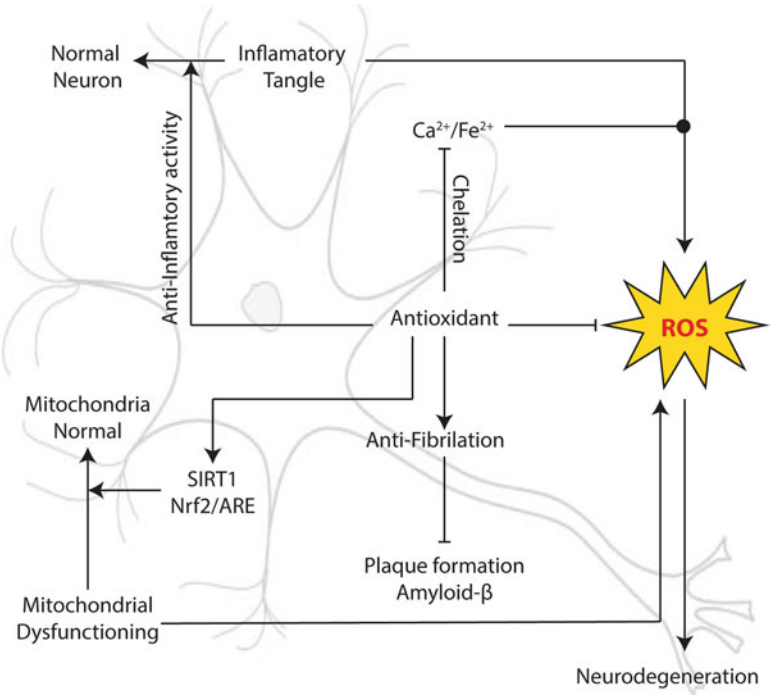


Fig. 13.2 Antioxidant strategies and sources of oxidative stress in AD

of these stimulate the synthesis of endogenous antioxidant molecules in cell via activation of Nrf/ARE pathways. Therefore, these compounds can be good candidates for the assessment of the AD by scavenging free radicals. Oxidative stress and its sources in AD are illustrated in (Fig. 13.2).

In our study we also took three cholinesterase inhibitors (donepezil, galantamine and rivastigmine), which are commonly prescribed drugs in Alzheimer, as reference compounds over our natural compounds for drug-likeness studies. These reference drugs provides mechanistic prevention for an enzyme called acetylcholinesterase, which breaks down acetylcholine in the brain and enhances antioxidants effects and

attenuates oxidative stresses (Klugman et al. 2012; Umukoro et al. 2014). As a result of our investigation, we found that our natural compounds show promising inhibitory activity too, while some of them were found to have even better activity than prescribed drugs against AD targets. Any compound cannot be directly considered as a drug molecule unless it is validated by several parameters like pharmacokinetic properties, ADME properties and potential toxicity. Therefore, with the help of various bioinformatics tool, we validated all our compounds.

Molecular docking studies are used to find out the interaction between a ligand/drug and a protein at the atomic level which allows us to characterize the behaviour of our compounds in the binding site of targets as well as to explain fundamental biochemical processes (Meng et al. 2011). Each of the natural compounds were docked with all 13 AD-associated proteins individually, to determine the best binding affinity using Autodock4.2 (Morris et al. 2009). Further, these compounds could be useful for the identification and development of new preventive and therapeutic drug against Alzheimer disease.

13.2 Materials and Method

13.2.1 Natural Compounds Selection

Forty-five natural compounds were selected as common reported antioxidant natural compounds from various database and literature. All chemical structures of these compounds were sketched in ChemBioDraw Ultra 12.0 (CambrigdeSoft) as shown in (Fig. 13.3).

13.2.2 Basic Pharmacokinetic Parameters Calculation

A compound has to be passed through multiple filters to be considered a novel drug. Most of the compounds that fail in preclinical trials do so because they do not show the required pharmacological properties to be a drug molecule (Wolohan and Clark 2003). Pharmacokinetic properties such as absorption, distribution, metabolism, excretion and toxicity (ADMET) have play a very crucial role in development of drug design to the final clinical success of a drug candidate (Moroy et al. 2012). Therefore, prediction of ADMET properties was done earlier with the aim of decreasing the failure rate of the compound for further process in future. Pharmacokinetic properties of natural compounds such as MW (molecular weight), LogP, HBD (number of hydrogen bond donors), HBA (number of hydrogen bond acceptors), TPSA (topological polar surface area), nrtB (number of rotatable bonds) and nViolation (violations of Lipinski's rule of

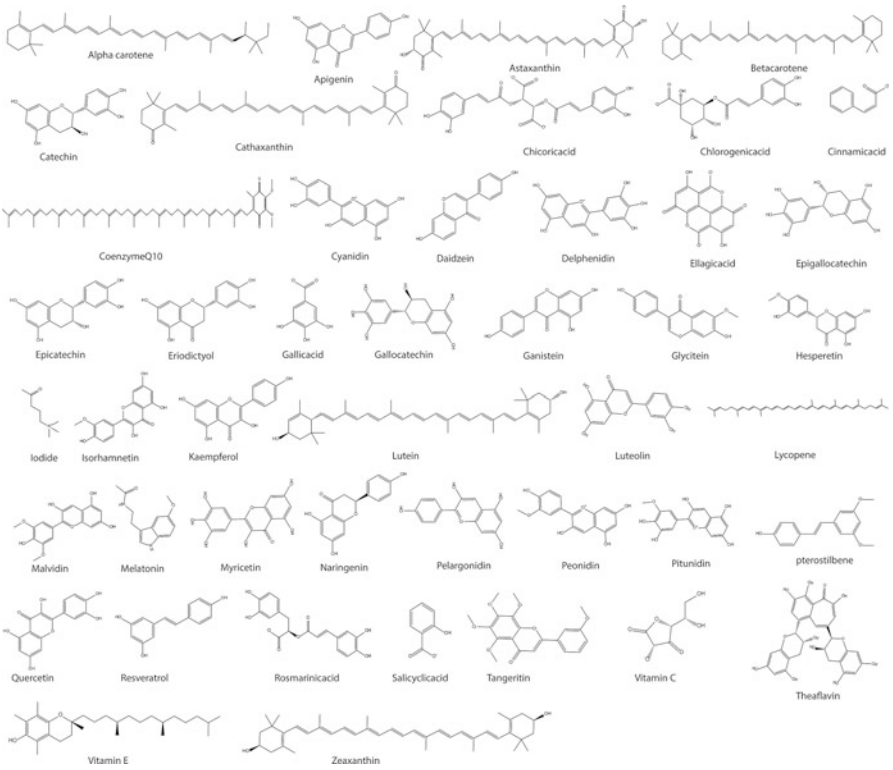


Fig. 13.3 Chemical structures of the studied compounds

five) (Lipinski et al. 2001) were calculated by DruLito (Drug LiknessTool) (www.niper.gov.in/pi_dev_tools/DruLiToWeb/DruLiTo_index.html) and Molinspiration Online tool (http://www.molinspiration.com/).

13.2.3 Compound Toxicity Prediction

The compounds toxicities prediction is an important sect of the drug design development process. In silico toxicity assessments are not only faster but can also reduce the amount of animal experiments. So, we calculated LD_{50} values for all our natural compounds, LD_{50} value is the amount of doses given to kill 50% of a test population (lab rats or other animals). It is an index determination of medicine and poison's virulence. **The lower the LD_{50} dose, the greater is the toxicity of the substance**. These LD_{50} values were calculated by an online tool ProTox (Drwal et al. 2014). We also determined its carcinogenic, mutagenic and skin irritation properties by using Discovery Studio 2.5 (Accelrys Software Inc., San Diego, CA, USA).

13.3 Molecular Docking

13.3.1 Target Preparation

All the 13 Alzheimer disease-associated targets were downloaded from Protein Data Bank (PDB) (http://www.rcsb.org/pdb/home/home.do) and are listed in Table 13.2. For crystal structure of each target, the crystallographic water molecules were removed, the missing hydrogen atoms were added, and the energy level of all the 13 targets was minimized using swiss_PDB viewer tool (Johansson et al. 2012).

13.3.2 Ligand Preparation

The structure of all the 45 compounds was drawn in **ChemDraw12** (PerkinElmer Informatics, Waltham, MA, USA) and converted to their 3D form, and also geometry of the compounds was optimized in **ChemBio3DUltra12** (PerkinElmer Informatics, Waltham, MA, USA). Finally all compounds were saved in PDB format for further docking studies.

13.3.3 Target Ligand Docking

Docking studies yielded crucial information concerning the orientation of the inhibitors in the binding pocket of the target proteins. During the molecular docking process, all the natural compounds bound in the groove of their respective targets. Each of the compounds was docked with all the 13 AD-associated targets; hence a total of 585 dockings were performed (Table 13.3). Then we used 3D sorting method to filter out the best possible compounds from the pool of 45 natural compounds.

Table 13.2 13 Alzheimer disease-associated targets

S.No	PDB ID	Protein name	References
1	1EQG	Cyclooxygenase-1 (COX-1)	Selinsky et al. (2001)
2	1MX1	Human carboxylesterase (hCE-1)	Bencharit et al. (2003)
3	1PBQ	N-methyl-D-aspartate (NMDA)	Furukawa and Gouaux (2003)
4	1Q5K	Glycogen-synthase-kinase-3β (GSK-3β)	Bhat et al. (2003)
5	1QWC	Nitric oxide synthase (NOS)	Fedorov et al. (2003)
6	1UDT	Phosphodiesterase-5 (PD-5)	Sung et al. (2003)
7	2FV5	TNF-α converting enzyme (TACE)	Niu et al. (2006)
8	3BKL	Angiotensin converting enzyme (ACE)	Watermeyer et al. (2008)
9	3G9N	c-Jun N-terminal kinase (JNK)	Cao et al. (2009)
10	3QMO	Cyclooxygenase-2 (COX-2)	Vecchio and Malkowski (2011)
11	4B0P	Butyrylcholinesterase (BuChE)	Wandhammer et al. (2013)
12	4DJU	β-site amyloid precursor protein	Cumming et al. (2012)
13	4EY5	Acetylcholinesterase (AChE)	Cheung et al. (2012)

Table 13.3 Binding energies of all the 45 natural compounds and 03 reference drugs docked with 13Alzheimer diseaseassociated targets

	1EQG	1MX1	1PBQ	1Q5K	1QWC	1UDT	2FV5	3BKL	3G9N	3QMO	4B0P	4DJU	4EY5
	Affinity (kcal. mol)	Affinity (kcal/ mol)	Affinity (kcal. mol)	Affinity (kcal/ mol)	Affinity (kcal/ mol)	Affinity (kcal/ mol)	Affinity (kcal/ mol)	Affinity (kcal/ mol)	Affinity (kcal/ mol)	Affinity (kcal/ mol)	Affinity (kcal/ mol)	Affinity (kcal/ mol)	Affinity (kcal/ mol)
Alpha carotene	−7.6	−7.3	−7.2	−8.6	−7.3	−7.5	−6.8	−7.6	−7.5	−7.4	−8	−8.8	−7.1
Apigenin	−9.1	**−9.3**	−8.4	−8.7	−9.3	−9.4	−8.0	−8.3	−7.8	−7.4	−7.9	−8.4	−8.9
Astaxanthin	−8.2	−7.9	−7.4	−8.7	−6.2	−7.0	−8.0	−10.0	−8.1	−7.4	−8.7	−9.4	−9.2
Beta carotene	−7.9	−7.1	**−8.7**	−7.4	−9.0	−7.7	−7.6	−7.8	−7.3	−7.9	−8.5	−9.3	−7.1
Canthaxanthin	−8.4	−7.9	−7.8	−8.1	−9.2	−7.2	−8.2	−7.6	**−8.7**	−7.5	−7.9	−7.4	−6.9
Catechin	−8.7	**−9.4**	**−8.5**	**−9.7**	**−9.9**	−7.7	**−10.1**	**−10.4**	**−9.0**	−7.6	**−10.6**	**−9.5**	**−9.6**
Chicoric acid	−8	−6.6	−7.3	−8.5	−8.6	−7.1	−8.5	−8.6	−7.4	−7.2	−9.6	−7.3	−8.2
chlorogenic acid	−7.6	−7.4	−6.9	−8.5	−4.8	−6.5	**−9.8**	−7.8	−7.5	−7.5	−8.3	−7.4	−7.6
Cinnamic acid	−5	−5.5	−4.2	−6.2	−4.7	−6.5	−4.7	−6.1	−4.7	−6.3	−5.4	−5.7	−4.8
Coenzyme Q10	−6.1	−4.8	−4.8	−7.5	−5.1	−6.7	−7.9	−6.8	−6.8	−5.6	−8.2	−7.1	−5.5
Cyanidin	**−9.4**	−7.2	−8.5	−8.4	−7.9	−7.7	**−9.6**	−8.9	−7.9	−7.6	−9.2	−8.2	−8.8
Daidzein	−8.5	−7.3	−7.8	−8.4	−7	−8.6	−9.3	−8.2	−7.7	−8.3	−8.3	−8.7	−8.2
Delphinidin	−8.1	−7.9	−7.6	−9.0	−8.6	−7.2	−9.6	−8.3	**−8.4**	**−9.2**	−9.3	−8.3	−8.6
Ellagic acid	−7.3	−7.5	−8.2	**−9.4**	−8.4	**−9.7**	−8.3	−8.3	−7.6	−8.1	**−10.5**	−8.1	**−9.7**
Epicatechin	−9	−7.4	−7.9	−8.7	**−9.7**	**−9.5**	−9.5	−8.3	−7.5	−8.9	−8.9	−8.2	−8.2
Epigallocatechin	−7.9	−7.5	−7.5	−8.1	−6.9	**−9.6**	−8.3	−8.3	−7.5	−9	−8.8	−7.7	−7.9
Eriodictyol	**−9.5**	−7.6	−7.8	−8.7	−7	**−9.5**	**−9.8**	−8.4	−7.6	**−9.4**	**−9.4**	−8.2	−8.3
Gallic acid	−6.2	−5.6	−5.8	−6	−6.2	−6.1	−6.9	−6.3	−5.8	−5.2	−6.4	−5.5	−6.4
Gallocatechin	−7.9	−7.5	−7.4	−8.2	−6.9	−9.6	−7.4	−8.2	−7.4	−7.6	−9.2	−7.8	−8.4
Genistein	−7.7	−7.4	**−8.7**	−8.8	**−9.6**	−7.2	−9	−8.3	−8	−7.8	−7.9	−9.1	−8.6
Glycitein	−7.4	**−8.1**	−7.5	**−9.1**	−6.8	−7.8	−7.9	**−9.4**	−7.2	−7.3	−8.1	**−10.2**	−9.1
Hesperetin	−8.2	−7.6	−8.3	−8.4	−7.1	−7.2	−9.2	−8.4	−7.9	**−9.2**	−9.1	−8.4	−8.4
Iodide	−4	−3.9	−4.6	−4.3	−4.7	−3.5	−5.1	−4.4	−4.3	−3.8	−4.5	−4.1	−3.5
Isorhamnetin	**−9.5**	−7	−7.6	−8.8	−7.3	−8.5	−8.9	−7.8	−8	−9	−9.2	−8.4	−7.3
Kaempferol	−7.7	−7.2	−7.4	−8.8	−6.8	−8.7	−9.0	−8.2	−8.2	−8.9	−7.8	−8.5	−8.9
Lutein	−8.6	**−8.4**	−6.5	−7.5	−7.8	−7.7	−9.3	−7.6	−8	−7.1	−8.6	−8.9	−7.3

(continued)

Table 13.3 (continued)

	1EQG	1MX1	1PBQ	1Q5K	1QWC	1UDT	2FV5	3BKL	3G9N	3QMO	4B0P	4DJU	4EY5
	Affinity (kcal. mol)	Affinity (kcal/ mol)	Affinity (kcal. mol)	Affinity (kcal/ mol)	Affinity (kcal/ mol)	Affinity (kcal/ mol)	Affinity (kcal/ mol)	Affinity (kcal/ mol)	Affinity (kcal/ mol)	Affinity (kcal/ mol)	Affinity (kcal/ mol)	Affinity (kcal/ mol)	Affinity (kcal/ mol)
Luteolin	−8	−7.6	**−8.6**	−8.7	−9.0	**−9.8**	**−9.7**	−8.5	−7.8	**−9.2**	**−9.7**	−8.4	−8.6
Lycopene	−6.4	−6.6	−6.6	−6.2	−1.9	−7.4	−8.0	−7.8	−6.8	−6.2	−6.8	−7.1	−6.2
Malvidin	−8.5	−6.8	−7.3	−8.5	−8.2	−7.5	−8.7	−8	−7.5	−8.8	−8.9	−8.1	−7.9
Melatonin	−7.5	−6.5	−6.9	−7.1	−5.5	−7.9	−6.1	−7.2	−6.4	−6.8	−7.9	−7	−7.2
Myricetin	−9.2	−7.1	−7.6	−8.9	−7.1	−6.5	−8	−8.3	−8	−8	−9.4	−7.9	−8.5
Naringenin	−8.5	−7.3	−7.5	−8.6	**−9.5**	**−9.5**	−9	−8.3	−7.6	−9	−9.1	−8.2	−8.5
Pelargonidin	−8.3	−8.1	−9.1	**−9.7**	**−10.1**	−6.9	−8	**−9.4**	−9	−7.9	**−10.1**	−9	**−10.3**
Peonidin	−7.1	−7.3	−7.6	−8.4	−9.2	−9	−9.4	−8.1	−7.6	**−9.2**	−9	−8.1	−8.9
Petunidin	−8	−7.1	−7.6	−9.0	−8.6	−9.4	−8.8	−8.2	**−8.4**	−7.6	−9.3	−8.2	−8
Pterostilbene	−7.2	−6.4	−7	−7.5	−6.5	−8.4	−7.8	−7.5	−6.2	−7.7	−7	−7.6	−7.5
Quercetin	**−9.5**	−7.1	−7.6	−8.5	−7.3	**−9.4**	−9.2	−8.5	**−8.7**	−8.4	−9.4	−8.2	−8.5
Resveratrol	−6.5	−6.5	−6.9	−7.3	−8.5	−6.4	−8.6	−7.2	−6.4	−6.5	−7.6	−7.4	−8
Rosmarinic acid	−7.9	−7.4	−5.8	−8.5	−7.8	−7.1	−8.3	−8.5	−5.9	−8.1	−9.2	−7.4	−8.2
Salicyclic acid	−5.5	−5.8	−5.7	−5.9	−4.9	−5.4	−6.2	−6.2	−5.2	−5.4	−5.7	−5.1	−6.3
Tangeritin	−6.8	−6.8	−7.8	−8	−6.8	−9	−8.2	−7.8	−7.6	−7.4	−7.1	−8.2	−8
Theaflavin	−9.8	−8.6	−10.8	−11.4	−8.9	−8	−11.2	−11.3	−8.9	−9.3	−10.2	−10.8	−11.1
Vitamin C	−5.1	−5.1	−5.3	−5.4	−5.8	−4.8	−6.7	−6.1	−5.1	−5.6	−5.8	−5.5	−5.3
Vitamin E	−6.8	−5.7	−7.3	−5.4	−4.7	−5.2	−4.1	−5	−5	−5.2	−5.9	−3.7	−5
Zeaxanthin	−7.8	−7.8	−7.5	−8	−8.2	−6.9	−7.8	−8.7	−8.3	−7.2	−8.4	**−9.6**	−7.9
Drugs commonly prescribed in Alzheimer disease													
Donepezil	**−8.9**	**−6.5**	**−7.5**	**−8.9**	**−9.8**	**−6.7**	**−6.4**	**−8.1**	**−8.6**	**−7.6**	**−7.7**	**−8.6**	**−7.5**
Galantamine	−7.4	−6.5	−5.7	−6.8	−8.0	−5.8	−6.4	−7.7	−6.7	−6.8	−6.8	−6.3	−6.5
Rivastigmine	−6.4	−6.1	−6.6	−6.7	−8.3	−5.2	−7.2	−6.1	−6.1	−5.9	−7.2	−6.2	−6.2

This method can be defined as filtering out the best using three levels of selection, first on the basis of the score (docking), second in addition to ranks the compounds on the basis of interaction with the number of receptors and finally eliminating those that violate the minimum criterion for RO5 (Lipinski Rule of Five) and toxicity. The docking interactions were visualized with PYMOL molecular graphics system, version 1.7.4.4 (Schrödinger, LLC, and Portland, OR, USA) and Maestro Visualizer (Maestro, Schrödinger, LLC, New York, NY, 2017). The docking studies were performed using Autodock4.0 (Morris et al. 2009). The inhibition constant (K_i) of natural compounds against Alzheimer-associated targets was calculated from docking energy using the following equation:

$$Ki = \exp(\Delta G * 1000)/RT)$$

where ΔG = docking energy; $R = 1.98719$ cal K^{-1} $mol^{-1;\ T\ 298.15^{\circ}K,\ Ki}$ = inhibition constant (nM).

13.4 Results

13.4.1 Pharmacokinetic Properties

Pharmacokinetic properties of natural compounds to be considered as drug candidates were based on Lipinski's rule of five. This rule is formulated for most orally administered drugs; it uses four criteria to determine if a molecule is drug-like: to have a molecular weight of $\leq$500, a LogP (logarithm of partition coefficient) $\leq$5, five or fewer hydrogen bond donor sites, and ten or fewer hydrogen bond acceptor sites. Molecules violating more than one of these rules may have problems with bioavailability. The entire set of compounds well followed the RO5 (rule of 5) except 15 of the compounds, out of which eight compounds (*alpha-carotene*, *astaxanthin*, *beta-carotene*, *canthaxanthin*, *coenzyme Q10*, *epigallocatechin*, *gallocatechin and theaflavin*) violated more than one of these rules by having *molecular mass* > 500, *logP* > 5 and $H - bond\ doner > 5$ that can create a problem in oral bioavailability.

TPSA analysis checked the bioavailability of natural compounds, as per the Veber's rule for good oral bioavailability, the number of rotatable bond must be $\leq$10 and TPSA values $\leq$140*Å* (Veber et al. 2002). The number of rotatable bonds has been shown to be a very good descriptor of oral bioavailability of drugs and has been found better to discriminate between compounds that have oral bioavailability of drugs. Rotatable bond is defined as any single non-ring bond, bounded to nonterminal heavy (i.e. non-hydrogen) atom. Amide C–N bonds are not considered because of their high-rotational energy barrier (Veber et al. 2002). The numbers of rotatable bonds in all of our compounds were found to be appropriate as in reference compounds (donepezil, galantamine and rivastigmine) except four compounds (*vit E*, *chicoric acid*, *coenzymeQ10*, *lycopene*) that have >10*nRB*. Once TPSA was calculated and it was found that only four compounds (*chicoric acid*, *chlorogenic acid*, *rosmarinic acid and theaflavin*) have TPSA values $\leq$140 Å, then we calculated percentage of absorption for all the 45 compounds using Zhao et al. formula (Zhao et al. 2002).

$$\text{Percentages of Absorption } (\%\text{ABS}) = [109 - (0.3345 \times TPSA)]$$

According to above formula, we calculated percentages of absorption of our compounds and our reference compounds. Compounds *(theaflavin*, *chicoric acid*, *chlorogenic acid*, *myricetin and rosmarinic acid)* have poor absorbance percentage of 36.212%, 37.490%, 52.944%, 59.601% and 59.711%, respectively; further details are given in (Table 13.4).

Table 13.4 Shows the drug-likeness of compounds and violation of Lipinski's rule are highlighted in red colour

S. No.	Ligand	Molecular Formula	Molecular Weight (g/mol)	logP	H-Bond Acceptors	H-Bond Donors	TPSA	% ABS	nRB	nAtom	nViolation
1	Alpha carotene	$C_{40}H_{56}$	536.44	15.187	0	0	0	109.00	10	96	2
2	Apigenin	$C_{15}H_{10}O_5$	270.05	1.567	1	3	86.99	79.901	1	30	0
3	Astaxanthin	$C_{40}H_{52}O_4$	596.39	9.696	4	2	74.6	84.046	10	96	2
4	Beta carotene	$C_{40}H_{56}$	536.44	14.734	0	0	0	109.00	10	96	2
5	Catechin	$C_{15}H_{14}O_6$	290.08	0.406	1	5	110.38	72.077	1	35	0
6	Canthaxanthin	$C_{40}H_{52}O_2$	564.4	10.788	2	0	34.14	97.580	10	94	2
7	Chicoric acid	$C_{22}H_{18}O_{12}$	472.06	0.722	8	4	213.78	37.490	11	50	0
8	Chlorogenic acid	$C_{16}H_{18}O_9$	353.09	-0.928	7	5	167.58	52.944	5	42	1
9	Cinnamic acid	$C_9H_8O_2$	147.04	3.423	2	0	40.13	95.576	2	18	0
10	Coenzyme Q10	$C_{59}H_{90}O_4$	862.68	18.454	4	0	52.6	91.405	31	153	2
11	Cyanidin	$C_{15}H_{11}O_6^+$	287.06	0	1	5	101.15	75.165	1	32	0
12	Daidzein	$C_{15}H_{10}O_4$	254.06	2.475	1	2	66.76	86.668	1	29	0
13	Delphinidin	$C_{15}H_{11}O_7^+$	303.05	0	1	6	121.38	68.398	1	33	1
14	Ellagic acid	$C_{14}H_6O_8$	299.99	0.652	8	2	93.06	77.871	0	26	0
15	Epicatechin	$C_{15}H_{14}O_6$	290.08	0.406	1	5	110.38	72.077	1	35	0
16	Epigallocatechin	$C_{15}H_{14}O_7$	306.07	-0.128	1	6	130.61	65.310	1	36	2
17	Eriodictyol	$C_{15}H_{12}O_6$	288.06	0.363	1	4	107.22	73.134	1	33	0
18	Gallic acid	$C_7H_6O_5$	169.01	-0.479	2	3	100.82	75.275	1	17	0
19	Gallocatechin	$C_{15}H_{14}O_7$	306.07	-0.128	1	6	130.61	65.310	1	36	1
20	Genistein	$C_{15}H_{10}O_5$	270.05	1.043	5	3	86.99	79.901	1	30	0
21	Glycitein	$C_{16}H_{12}O_5$	284.07	2.031	1	2	75.99	83.581	2	33	0
22	Hesperetin	$C_{16}H_{14}O_6$	302.08	0.453	1	3	96.22	76.814	2	36	0
23	Iodide	I^-	144.14	0.035	1	0	17.07	103.29	4	28	0
24	Isorhamnetin	$C_{16}H_{12}O_7$	316.06	1.471	2	4	116.45	70.047	2	35	0
25	Kaempferol	$C_{15}H_{10}O_6$	286.05	1.915	2	4	107.22	73.134	1	31	0
26	Lutein	$C_{40}H_{56}O_2$	568.43	11.283	2	2	40.46	95.466	10	98	1
27	Luteolin	$C_{15}H_{10}O_6$	286.05	1.033	1	4	107.22	73.134	1	31	0
28	Lycopene	$C_{40}H_{56}$	536.44	14.586	0	0	0	109.00	16	96	1
29	Malvidin	$C_{17}H_{13}O_7+$	331.08	0	1	4	99.38	75.757	3	39	0
30	Melatonin	$C_{13}H_{16}N_2O_2$	232.12	1.254	3	2	50.36	92.154	5	33	0
31	Myricetin	$C_{15}H_{10}O_8$	318.04	0.847	2	6	147.68	59.601	1	33	1
32	Naringenin	$C_{15}H_{12}O_5$	272.07	0.897	1	3	86.99	79.901	1	32	0
33	Pelargonidin	$C_{15}H_{11}O_5^+$	271.06	0	1	4	80.92	81.932	1	31	0
34	Peonidin	$C_{16}H_{13}O_6^+$	301.07	0	6	4	90.15	78.844	2	35	0
35	Petunidin	$C_{16}H_{13}O_7^+$ (Cl^-)	317.07	0	1	5	110.38	72.077	2	36	0
36	Pterostilbene	$C_{16}H_{16}O_3$	256.11	2.69	3	1	38.69	96.058	4	35	0
37	Quercetin	$C_{15}H_{10}O_7$	302.04	1.381	2	5	127.45	66.367	1	32	0
38	Resveratrol	$C_{14}H_{12}O_3$	228.08	3.436	0	3	60.69	88.699	2	29	0
39	Rosmarinic acid	$C_{18}H_{16}O_8$	359.08	1.602	4	4	147.35	59.711	7	41	0
40	Salicyclic acid	$C_7H_6O_3$	137.02	0.8	2	1	60.36	88.809	1	15	0
41	Tangeritin	$C_{20}H_{20}O_7$	372.12	2.236	1	0	72.45	84.765	6	47	0
42	Theaflavin	$C_{29}H_{24}O_{12}$	564.13	-0.216	4	9	217.6	36.212	2	65	2
43	Vitamin C	$C_6H_8O_6$	175.02	-1.813	6	2	83.83	80.958	2	19	0
44	Vitamin E	$C_{29}H_{50}O_2$	430.38	10.938	0	1	29.46	99.145	12	81	1
45	Zeaxanthin	$C_{40}H_{56}O_2$	568.43	10.564	2	2	40.46	95.466	10	98	1
	Three Cholinesterase inhibitors are commonly prescribed in Alzheimer disease										
1	Donepezil	$C_{24}H_{30}NO_3^+$	380.50	3.65	1	4	39.97	95.23	6	58	0
2	Galantamine	$C_{17}H_{12}NO_3^+$	288.36	1.54	2	4	43.13	94.14	1	43	0
3	Rivastigmine	$C_{14}H_{23}N_2O_2^+$	251.34	2.28	1	4	33.98	97.27	6	41	0

%ABS- Percentage of Absorption **TPSA-** Topological Polar Surface Area **nrtB-** Number of rotable Atoms
nAtom- Number of Atoms **nViolation-** Violation of Lipinski's rule **XlogP d** 5
H-BD < 5 **H-BA** < 10 **MW** < 500

%ABS percentage of absorption, *TPSA* topological polar surface area, *nrtB* number of rotable atoms, *nAtom* number of atoms, *nViolation* violation of Lipinski's rule, XlogP ≤5, H-BD <5, H-BA <10, MW <500

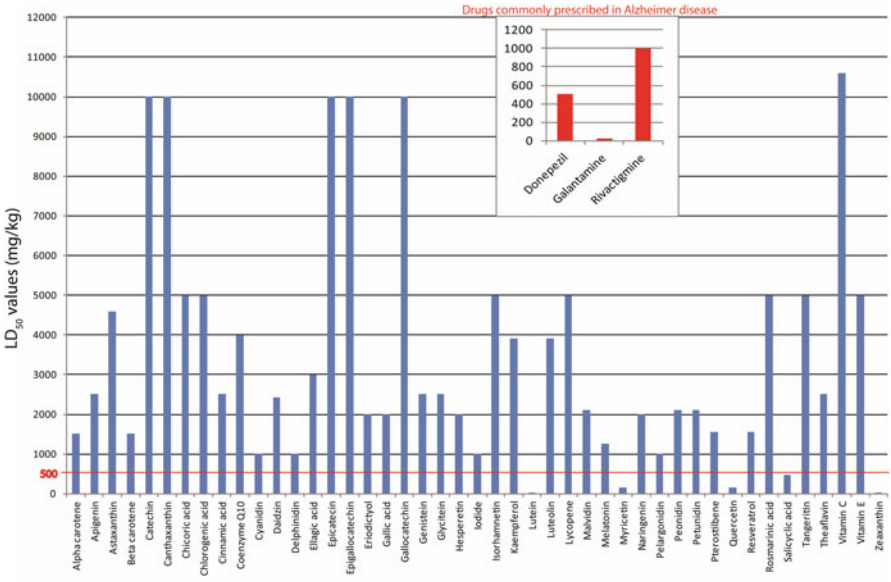


Fig. 13.4 Histogram representation of *LD*50 values of natural compounds compared with commonly prescribes drugs

13.4.2 Toxicity Prediction

In the computational prediction of toxicities, drug score profiles of natural compounds are promising. An online software PROTOX was used for the prediction of the LD_{50} of the new compounds. Most of the compounds in our study fell in non-toxic zone (above the 1000 mg/kg) except lutein (10 mg/kg), myricetin (159 mg/kg), quercetin (159 mg/kg), salicylic acid (480 mg/kg) and zeaxanthin (10 mg/kg); results are shown in the graph in Fig. 13.4. The mutagenicity, carcinogenicity and skin irritation properties were also predicted, and it was found that 05 compounds are mutagenic, 07 compounds are carcinogenic, 38 compounds are very sensitive to skin, but 09 of these compounds may cause severe skin irritation problems. Results are listed in (Table 13.5).

13.4.3 Molecular Docking Studies

To ensure the interaction between the natural compounds and Alzheimer disease-associated targets, we performed molecular docking analysis using Autodock4.2. Each of the compounds was docked with all the 13 Alzheimer disease-associated targets individually. These compounds showed very good binding affinity with all the 13 Alzheimer-associated targets.

Table 13.5 Shows the toxicity of compounds and various toxicitiness that are highlighted in red colour

S. no.	Compound	Carcinogenicity	Mutagenicity	Skin Irritation	Skin Sensitization	Biodegradability
1	Alpha carotene	Carcinogen	Non mutagen	Severe	Weak	Degradable
2	Apigenin	Non carcinogen	Non mutagen	None	Strong	Non degradable
3	Astaxanthin	Non carcinogen	Non mutagen	Mild	Weak	Degradable
4	Beta carotene	Carcinogen	Non mutagen	Severe	Weak	Degradable
5	Catechin	Non carcinogen	Non mutagen	None	Strong	Non degradable
6	Canthaxanthin	Carcinogen	Non mutagen	Severe	Weak	Degradable
7	Chicoric acid	Non carcinogen	Non mutagen	None	Strong	Degradable
8	Chlorogenic acid	Non carcinogen	Non mutagen	Mild	None	Degradable
9	Cinnamic acid	Carcinogen	Non mutagen	Mild	Strong	Degradable
10	Coenzyme Q10	Carcinogen	Non mutagen	Severe	None	Degradable
11	Cyanidin	Non carcinogen	Non mutagen	None	Strong	Non degradable
12	Daidzein	Non carcinogen	Non mutagen	None	Strong	Non degradable
13	Delphinidin	Non carcinogen	Non mutagen	None	Strong	Non degradable
14	Ellagic acid	Non carcinogen	Non mutagen	Mild	Weak	Non degradable
15	Epicatechin	Non carcinogen	Non mutagen	None	Strong	Non degradable
16	Epigallocatechin	Non carcinogen	Non mutagen	None	Strong	Non degradable
17	Eriodictyol	Non carcinogen	Non mutagen	None	Strong	Non degradable
18	Gallic acid	Non carcinogen	Non mutagen	None	Strong	Degradable
19	Gallocatechin	Non carcinogen	Non mutagen	None	Strong	Non degradable
20	Genistein	Non carcinogen	Non mutagen	None	Strong	Non degradable
21	Glycitein	Non carcinogen	Non mutagen	None	None	Degradable
22	Hesperetin	Non carcinogen	Non mutagen	None	Strong	Degradable
23	Iodide	Non carcinogen	Non mutagen	Severe	Weak	Degradable
24	Isorhamnetin	Non carcinogen	Mutagen	None	Strong	Non degradable
25	Kaempferol	Non carcinogen	Mutagen	None	Strong	Non degradable
26	Lutein	Non carcinogen	Non mutagen	Severe	Weak	Degradable
27	Luteolin	Non carcinogen	Non mutagen	None	Strong	Non degradable
28	Lycopene	Carcinogen	Non mutagen	Severe	Weak	Degradable
29	Malvidin	Non carcinogen	Non mutagen	None	Strong	Non degradable
30	Melatonin	Non carcinogen	Non mutagen	None	None	Non degradable
31	Myricetin	Non carcinogen	Mutagen	None	Strong	Non degradable
32	Naringenin	Non carcinogen	Non mutagen	None	Strong	Non degradable
33	Pelargonidin	Non carcinogen	Non mutagen	None	Strong	Non degradable
34	Peonidin	Non carcinogen	Non mutagen	None	Strong	Non degradable
35	Petunidin	Non carcinogen	Non mutagen	None	Strong	Non degradable
36	Pterostilbene	Non carcinogen	Mutagen	None	Strong	Non degradable
37	Quercetin	Non carcinogen	Non mutagen	None	Strong	Non degradable
38	Resveratrol	Non carcinogen	Non mutagen	None	Strong	Non degradable
39	Rosmarinic acid	Non carcinogen	Non mutagen	None	Strong	Non degradable
40	Salicyclic acid	Non carcinogen	Non mutagen	None	None	Degradable
41	Tangeritin	Non carcinogen	Non mutagen	None	Strong	Degradable
42	Theaflavin	Non carcinogen	Non mutagen	None	Strong	Non degradable
43	Vitamin C	Non carcinogen	Non mutagen	None	None	Degradable
44	Vitamin E	Carcinogen	Non mutagen	Severe	None	Non degradable
45	Zeaxanthin	Non carcinogen	Non mutagen	Severe	Weak	Degradable

Among all of the compounds, only 09 compounds interacted with maximum number of targets, such as *theaflavin* showing maximum interactions, as it interacted with 11 Alzheimer-associated targets (1*EQG*, 1*MX*1, 1*PBQ*, 1*Q*5*K*, 2*FV*5, 3*BKL*, 3*G*9*N*, 3*QMO*, 4*B*0*P*, 4*DJU* and 4*EY*5), whereas compound *catechin* interacted with 10 targets (1*MX*1, 1*PBQ*, 1*Q*5*K*, 1*QWC*, 2*FV*5, 3*BKL*, 3*G*9*N*, 4*B*0*P*, 4*DJU and* 4*EY*5), *pelargonidin* interacted with 05 targets (1*Q*5*K*, 1*QWC*, 3*BKL*, 4*B*0*P* and 4*EY*5), *luteolin* intercalated with 05 targets (1*PBQ*, 1*UDT*, 2*FV*5, 3*QMO* and 4*B*0*P*), *ellagic acid* compound interacted with 04 targets (1*Q*5*K*, 1*UDT*,

4*B0P* and 4*EY*5), *eriodictyol* intercalated with 04 targets (1*EQG*, 1*UDT*, 2*FV*5 and 3*QMO*), *glycitein* intercalated with 04 targets (1*MX*1, 1*Q5K*, 3*BKL* and 4*DJU*), *quercetin* intercalated with 03 targets (1*EQG*, 1*UDT* and 3*G9N*), and *astaxanthin* intercalated with 03 targets (3*BKL*, 4*DJU* and 4*EY*5). Even though compounds *astaxanthin* and *theaflavin* showed very good binding affinity, we did not consider them for further docking analysis because they did not comply with the Lipinski's rule. *Astaxanthin* has greater molecular mass (596.39 Da) and greater log P (9.696) value, whereas *theaflavin* has greater molecular mass (564.13 g/mol) and more than five H-bond donor. The range of the binding affinities of all the compounds lies between −4.00 and −11.4 *kcal/mol*. In addition, the residues surrounding the ligands (within 4 Å) in all docking results provide short-range polar interactions that stabilize the formation of complex (with the help of delocalization of charges on ligands). It is not possible to show all the binding poses, but we have shown 07 poses which have maximum binding affinity among all of them (Fig. 13.5). The inhibitor constant (K_i) is an indicator of the effectiveness of an inhibitor, with a greatly potent inhibitor being indicated by a low K_i because the smaller the K_i, the greater the binding affinity, and the smaller amount of medication is needed in order to inhibit the activity of that enzyme Sung et al. (2003). All of our compounds were found to have a very small *Ki* value except compound *cinnamic acid* (with *1EQG*, *1PBQ*, *1QWC*, *2FV5*, *3G9N*, *4B0P and 4EY5*), *chlorogenic acid* (*with* 1*QWC*), *coenzyme Q*10 (with *1MX1*, *1PBQ* and *1QWC)*, *gallic acid* (with *3QMO*), iodide (with all targets), *lycopene* (with *1QWC*), *salicylic acid* (with *1QWC*, *1UDT*, *3G9N*, *3QMO and 4DJU*), *vitamin C* (with *1EQG*, *1MX1*, *1PBQ*, *1Q5K*, *1UDT*, *3G9N* and *4EY5*), *vitamin E* (with *1Q5K*, *1QWC*, *1UDT*, *2FV5*, *3BKL*, *3G9N*, *3QMO*, *4DJU* and *4EY5*) and all three drugs commonly prescribed in Alzheimer disease (*donepezil*, *galantamine* and *rivastigmine*) that have very poor K_i values with *4DJU* (Table 13.6). In this way, all of these results suggest that our natural compounds potentially interfere with Alzheimer-associated targets, which should prompt further investigations to expose the mechanism of our compounds against Alzheimer disease in vivo.

13.5 Discussion

Natural products have been used since ancient times and are well recognized as sources of drugs in several human ailments. The healing ability of these herbs and medicinal plants draw attention to study natural products as a potentially valuable resource of drug molecules; they are evolutionarily optimized as drug-like molecules and remain the best sources of drugs and drug leads (Ansari and Khodagholi 2013; Newman and Cragg 2012). In our study, we chose 45 natural compounds that have remarkable antioxidant property and act mainly by scavenging free radical species. In recent years, significant data have been gathered, indicating the level of oxidative stress increase in the brain in AD condition. This may have a role in the pathogenesis of neuron degeneration and death (Badrul and Ekramu n.d.). Thus, protection and

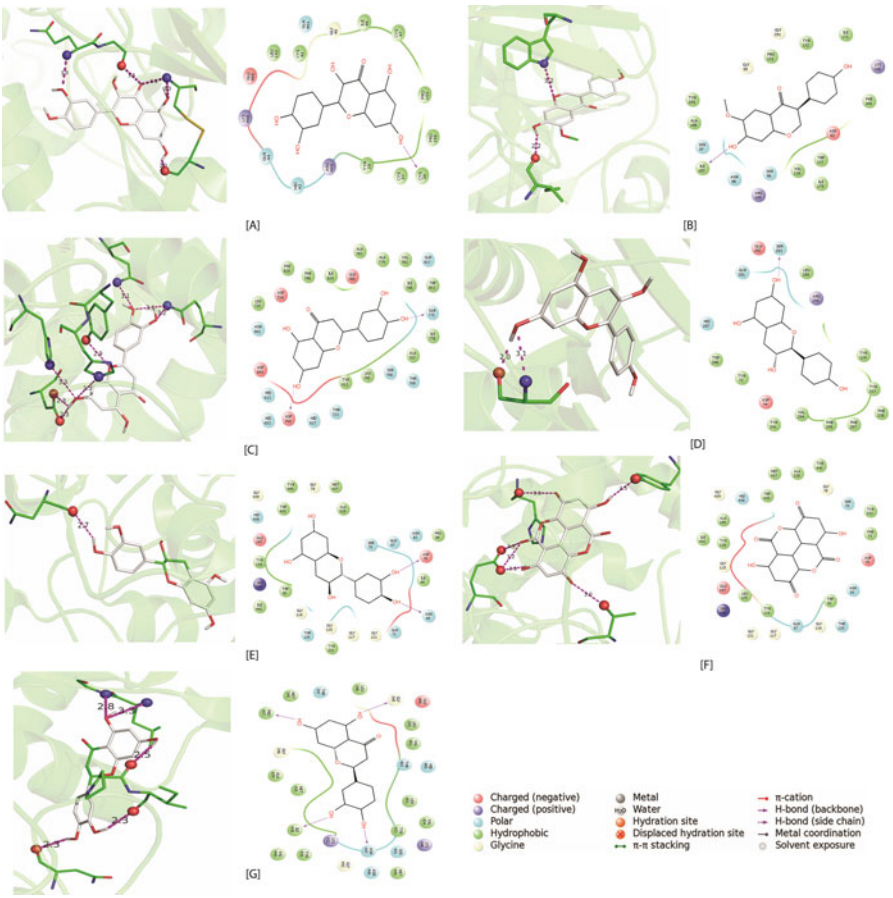


Fig. 13.5 Molecular docking poses of the top seven representative compounds: crystallographic structure of (**a**) *catechin* complexed to *butyrylcholinesterase(BuChE)(*Pdb id:4BOP), (**b**) *ellagic* acid complexed to *butyrylcholinesterase (BuChE)* (Pdb id: 4BOP), (**c**) *eriodictyol* complexed to *TNF-α converting enzyme (TACE)*(Pdb id: 2FV5), (**d**) *glycitein* complexed to *β-Site amyloid precursor protein* (Pdb id: 4DJU), (**e**) *luteolin* complexed to *TNF-α converting enzyme* (Pdb id: 2FV5), (**f**) *pelagonidin* complexed to *acetylcholinesterase (AChE)* (Pdb id: 4EY5) and (**g**) *quercetin* complexed to *cyclooxygenase-1 (COX-1)* (Pdb id: 1EQG)

inhibition against oxidative stress may be considered an important criterion in the development of anti-Alzheimer agents. Therefore, treatment with these antioxidant agents might prevent or reduce the progression of AD. These natural compounds may prove to be novel anti-Alzheimer agents. In order to exploit all the properties for a compound to behave as a drug, the study was done in silico by using different computational tools based on chemoinformatics or bioinformatics. The drug-like property predictions showed that most of the compounds followed the Lipinski's rule of five and ADMET. Finally, it can be seen from our docking studies also that the compounds exhibit a strong interaction with Alzheimer-associated targets. As

Table 13.6 Calculated *Ki*(*nM*) value from docking energy of 45 natural compounds against 13 Alzheimer disease-associated targets. Compounds that have very poor *Ki* values are highlighted in grey colour

Compound Names	1EQG	1MX1	1PBQ	1Q5K	1QWC	1UDT	2FV5	3BKL	3G9N	3QMO	4B0P	4DJU	4EY5
Alpha carotene	4.46E-06	4.46E-06	5.28E-06	4.97E-07	4.46E-06	3.18E-06	1.04E-05	2.69E-06	3.18E-06	3.76E-06	1.37E-06	3.54E-07	6.25E-06
Apigenin	2.14E-07	1.52E-07	6.96E-07	4.20E-07	1.52E-07	1.29E-07	1.37E-06	8.24E-07	1.92E-06	3.76E-06	1.62E-06	6.96E-07	2.99E-07
Astaxanthin	9.76E-07	1.62E-06	3.76E-06	4.20E-07	2.85E-05	7.39E-06	1.37E-06	4.68E-08	1.16E-06	3.76E-06	4.20E-07	1.29E-07	1.80E-07
Beta carotene	1.62E-06	6.25E-06	4.20E-07	3.76E-06	2.53E-07	2.27E-06	2.69E-06	1.92E-06	4.46E-06	1.62E-06	5.88E-07	1.52E-07	6.25E-06
Canthaxanthin	6.96E-07	1.62E-06	1.92E-06	1.16E-06	1.80E-07	5.28E-06	9.76E-07	2.69E-06	4.20E-07	3.18E-06	1.62E-06	3.76E-06	8.75E-06
Catechin	4.20E-07	1.29E-07	5.88E-07	7.76E-08	5.54E-08	2.27E-06	3.95E-08	2.38E-08	2.53E-07	2.69E-06	**1.70E-08**	1.09E-07	9.19E-08
Chicoric acid	1.37E-06	1.45E-05	4.46E-06	5.88E-07	4.97E-07	6.25E-06	5.88E-07	4.97E-07	3.76E-06	5.28E-06	9.19E-08	4.46E-06	9.76E-07
Chlorogenic acid	2.69E-06	3.76E-06	8.75E-06	5.88E-07	0.000303	1.72E-05	6.55E-08	1.92E-06	3.18E-06	3.18E-06	8.24E-07	3.76E-06	2.69E-06
Cinnamic acid	0.00021625	9.30E-05	0.000834	2.85E-05	0.000359	1.72E-05	0.000359	3.38E-05	0.000359	2.41E-05	0.00011	6.64E-05	0.000303
Coenzyme Q10	3.38E-05	0.000303	0.000303	3.18E-06	0.000183	1.23E-05	1.62E-06	1.04E-05	1.04E-05	7.85E-05	9.76E-07	6.25E-06	9.30E-05
Cyanidin	1.29E-07	5.28E-06	5.88E-07	6.96E-07	1.62E-06	2.27E-06	9.19E-08	2.99E-07	1.62E-06	2.69E-06	1.80E-07	9.76E-07	3.54E-07
Daidzein	5.88E-07	4.46E-06	1.92E-06	6.96E-07	7.39E-06	4.97E-07	1.52E-07	9.76E-07	2.27E-06	8.24E-07	8.24E-07	4.20E-07	9.76E-07
Delphinidin	1.16E-06	1.62E-06	2.69E-06	2.53E-07	4.97E-07	5.28E-06	9.19E-08	8.24E-07	6.96E-07	1.80E-07	1.52E-07	8.24E-07	4.97E-07
Ellagic acid	4.46E-06	3.18E-06	9.76E-07	1.29E-07	6.96E-07	7.76E-08	8.24E-07	8.24E-07	2.69E-06	1.16E-06	**2.01 E-08**	1.16E-06	7.76E-08
Epicatechin	2.53E-07	3.76E-06	1.62E-06	4.20E-07	7.76E-08	1.09E-07	1.09E-07	8.24E-07	3.18E-06	2.99E-07	2.99E-07	9.76E-07	9.76E-07
Epigallocatechin	1.62E-06	3.18E-06	3.18E-06	1.16E-06	8.75E-06	9.19E-08	8.24E-07	8.24E-07	3.18E-06	2.53E-07	3.54E-07	2.27E-06	1.62E-06
Eriodictyol	1.09E-07	2.69E-06	1.92E-06	4.20E-07	7.39E-06	1.09E-07	**6.55E-08**	6.96E-07	2.69E-06	1.29E-07	1.29E-07	9.76E-07	8.24E-07
Gallic acid	2.85E-05	7.85E-05	5.60E-05	4.00E-05	2.85E-05	3.38E-05	8.75E-06	2.41E-05	5.60E-05	0.000154	2.04E-05	9.30E-05	2.04E-05
Gallocatechin	1.62E-06	3.18E-06	3.76E-06	9.76E-07	8.75E-06	9.19E-08	3.76E-06	9.76E-07	3.76E-06	2.69E-06	1.80E-07	1.92E-06	6.96E-07
Genistein	2.27E-06	3.76E-06	4.20E-07	3.54E-07	9.19E-08	5.28E-06	2.53E-07	8.24E-07	1.37E-06	1.92E-06	1.62E-06	2.14E-07	4.97E-07
Glycitein	3.76E-06	1.16E-06	3.18E-06	2.14E-07	1.04E-05	1.92E-06	1.62E-06	1.29E-07	5.28E-06	4.46E-06	1.16E-06	**3.34E-08**	2.14E-07
Hesperetin	9.76E-07	2.69E-06	8.24E-07	6.96E-07	6.25E-06	5.28E-06	1.80E-07	6.96E-07	1.62E-06	1.80E-07	2.14E-07	6.96E-07	6.96E-07
Iodide	0.00116939	0.001384	0.000425	0.000705	0.000359	0.002719	0.000183	0.000595	0.000705	0.001639	0.000503	0.000988	0.002719
Isorhamnetin	1.09E-07	7.39E-06	2.69E-06	3.54E-07	4.46E-06	5.88E-07	2.99E-07	1.92E-06	1.37E-06	2.53E-07	1.80E-07	6.96E-07	4.46E-06
Kaempferol	2.27E-06	5.28E-06	3.76E-06	3.54E-07	1.04E-05	4.20E-07	2.53E-07	9.76E-07	9.76E-07	2.99E-07	1.92E-06	5.88E-07	2.99E-07
Lutein	4.97E-07	6.96E-07	1.72E-05	3.18E-06	1.92E-06	2.27E-06	1.52E-07	2.69E-06	1.37E-06	6.25E-06	4.97E-07	2.99E-07	4.46E-06
Luteolin	1.37E-06	2.69E-06	4.97E-07	4.20E-07	2.53E-07	**6.55E-08**	7.76E-08	5.88E-07	1.92E-06	1.80E-07	7.76E-08	6.96E-07	4.97E-07
Lycopene	2.04E-05	1.45E-05	1.45E-05	2.85E-05	0.040484	3.76E-06	1.37E-06	1.92E-06	1.04E-05	2.85E-05	1.04E-05	6.25E-06	2.85E-05
Malvidin	5.88E-07	1.04E-05	4.46E-06	5.88E-07	9.76E-07	3.18E-06	4.20E-07	1.37E-06	3.18E-06	3.54E-07	2.99E-07	1.16E-06	1.62E-06
Melatonin	3.18E-06	1.72E-05	8.75E-06	6.25E-06	9.30E-05	1.62E-06	3.38E-05	5.28E-06	2.04E-05	1.04E-05	1.62E-06	7.39E-06	5.28E-06
Myricetin	1.80E-07	6.25E-06	2.69E-06	2.99E-07	6.25E-06	1.72E-05	1.37E-06	8.24E-07	1.37E-06	1.37E-06	1.29E-07	1.62E-06	5.88E-07
Naringenin	5.88E-07	4.46E-06	3.18E-06	4.97E-07	1.09E-07	1.09E-07	2.53E-07	8.24E-07	2.69E-06	2.53E-07	2.14E-07	9.76E-07	5.88E-07
Pelargonidin	8.24E-07	1.16E-06	2.14E-07	7.76E-08	3.95E-08	8.75E-06	1.37E-06	1.29E-07	2.53E-07	1.62E-06	3.95E-08	2.53E-07	**2.82E-08**
Peonidin	6.25E-06	4.46E-06	2.69E-06	6.96E-07	1.80E-07	2.53E-07	1.29E-07	1.16E-06	2.69E-06	1.80E-07	2.53E-07	1.16E-06	2.99E-07
Petunidin	1.37E-06	6.25E-06	2.69E-06	2.53E-07	4.97E-07	1.29E-07	3.54E-07	9.76E-07	6.96E-07	2.69E-06	1.52E-07	9.76E-07	1.37E-06
Pterostilbene	5.28E-06	2.04E-05	7.39E-06	3.18E-06	1.72E-05	6.96E-07	1.92E-06	3.18E-06	2.85E-05	2.27E-06	7.39E-06	2.69E-06	3.18E-06
Quercetin	**1.09E-07**	6.25E-06	2.69E-06	5.88E-07	4.46E-06	**1.09E-07**	1.80E-07	5.88E-07	4.20E-07	6.96E-07	1.29E-07	9.76E-07	5.88E-07
Resveratrol	1.72E-05	1.72E-05	8.75E-06	4.46E-06	5.88E-07	2.04E-05	4.97E-07	5.28E-06	2.04E-05	1.72E-05	2.69E-06	3.76E-06	1.37E-06
Rosmarinic acid	1.62E-06	3.76E-06	5.60E-05	5.88E-07	1.92E-06	6.25E-06	8.24E-07	5.88E-07	4.73E-05	1.16E-06	1.80E-07	3.76E-06	9.76E-07
Salicyclic acid	9.30E-05	5.60E-05	6.64E-05	4.73E-05	0.000256	0.00011	2.85E-05	2.85E-05	0.000154	0.00011	6.64E-05	0.000183	2.41E-05
Tangeritin	1.04E-05	1.04E-05	1.92E-06	1.37E-06	1.04E-05	2.53E-07	9.76E-07	1.92E-06	2.69E-06	3.76E-06	6.25E-06	9.76E-07	1.37E-06
Theaflavin	6.55E-08	4.97E-07	1.21E-08	4.40E-09	2.99E-07	1.37E-06	6.17E-09	5.21E-09	2.99E-07	1.52E-07	3.34E-08	1.21E-08	7.30E-09
Vitamin C	0.00018266	0.000183	0.00013	0.00011	5.60E-05	0.000303	1.23E-05	3.38E-05	0.000183	7.85E-05	5.60E-05	9.30E-05	0.00013
Vitamin E	1.04E-05	6.64E-05	4.46E-06	0.00011	0.000359	0.000154	0.000988	0.000216	0.000216	0.000154	4.73E-05	0.00194	0.000216
Zeaxanthin	1.92E-06	1.92E-06	3.18E-06	1.37E-06	9.76E-07	8.75E-06	1.92E-06	4.20E-07	8.24E-07	5.28E-06	6.96E-07	9.19E-08	1.62E-06
Drugs commonly prescribed in Alzheimer disease													
Donepezil	3.18E-06	4.46E-06	5.28E-06	4.97E-07	4.46E-06	3.18E-06	1.04E-05	2.69E-06	3.18E-06	3.76E-06	1.37E-06	1.000001	6.25E-06
Galantamine	1.29E-07	1.52E-07	6.96E-07	4.20E-07	1.52E-07	1.29E-07	1.37E-06	8.24E-07	1.92E-06	3.76E-06	1.62E-06	1.000001	2.99E-07
Rivastigmine	7.39E-06	1.62E-06	3.76E-06	4.20E-07	2.85E-05	7.39E-06	1.37E-06	4.68E-08	1.16E-06	3.76E-06	4.20E-07	1.000000	1.80E-07

this is an in silico study, therefore one should not forget to consider the complex biological metabolic processes during the onset of Alzheimer. The interaction parameter may vary due to multiple interactions in complex system which is another subject of consideration. Molecular docking and pharmacokinetic studies showed that most of our compounds fulfil the requirements for an anti-Alzheimer drug, such as ADMET, RO5, non-toxicity, binding affinity, inhibition constants, antioxidant and neuroprotective inhibitory properties and good interaction with Alzheimer associated targets. We found 7 compounds (Catechin, Pelargonidin, Luteolin, Ellagic acid, Eriodictyol, Glycitein and Quercetin) may be considered to be a kind of novel

anti-Alzheimer disease drug agents, which are multi-target-directed ligands with not only antioxidant activity but also inhibitory and neuroprotective activities. In this way these results suggest that natural compounds potentially interfere with Alzheimer-associated targets, which should prompt further investigations to expose the mechanism of our compounds against Alzheimer disease in vivo. Therefore, we recommend them for future in vivo studies and possible clinical trials.

13.6 Conclusion

Several natural products are used alone or in combination with other neuroprotective compounds to improve memory and cognition in AD patients as supported by various experimental studies. Altogether, this pioneering study was used to preliminarily investigate the potential compounds from natural products and conventional docking study to analyse the best docked ligands permitted us to know the binding mode of compounds, binding energies of the drug-targets interactions are important to describe how fit the drug binds to the target macromolecule. Our study extends the knowledge on natural compounds showing its potential in curing Alzheimer disease. Most importantly, our findings introduce seven natural therapeutic agents for Alzheimer disease, which are multi target-directed ligands with antioxidant, inhibitory and neuroprotective activities. Therefore, these compounds may have potential in treatment for Alzheimer and other neurodegenerative diseases. Further studies are essential to explore the target-specific effect of these natural compounds on various signalling pathways, mode of action in various brain regions, the ability to cross the blood-brain barrier and the mechanism behind the synergistic action of the antioxidant agents on the target. Using novel pharmaceutical engineering and medicinal chemistry approach to prepare novel formulations or design, new compounds based on natural templates opens up a new window into using natural therapeutic agents against AD.

Acknowledgements The authors are grateful to the Centre for Interdisciplinary Research in Basic Sciences (CIRBSc), Jamia Millia Islamia for providing the research infrastructure.

Aftab Alam is supported by a research fellowship of the University Grants Commission, Government of India; SA and MZM are supported by the Indian Council of Medical Research under SRF (Senior Research Fellowship).

Disclosure The authors report no conflicts of interest in this work.

References

Ansari N, Khodagholi F (2013) Natural products as promising drug candidates for the treatment of Alzheimer's disease: molecular mechanism aspect. Curr Neuropharmacol 11(4):414–429. https://doi.org/10.2174/1570159X11311040005

Badrul A, Ekramu H (n.d.) Anti-Alzheimer and antioxidant activity of celastrus paniculatus seed. Iran J Pharm Sci 7(1):49–56

Bencharit S, Morton CL, Hyatt JL, Kuhn P, Danks MK, Potter PM, Redinbo MR (2003) Crystal structure of human carboxylesterase 1 complexed with the Alzheimer's drug tacrine: from binding promiscuity to selective inhibition. Chem Biol 10(4):341–349

Bhat R, Xue Y, Berg S, Hellberg S, Ormo M, Nilsson Y, . . . Avila J (2003) Structural insights and biological effects of glycogen synthase kinase 3-specific inhibitor AR-A014418. J Biol Chem 278(46):45937–45945. https://doi.org/10.1074/jbc.M306268200

Cao J, Gao H, Bemis G, Salituro F, Ledeboer M, Harrington E, . . . Green J (2009) Structure-based design and parallel synthesis of N-benzyl isatin oximes as JNK3 MAP kinase inhibitors. Bioorg Med Chem Lett 19(10):2891–2895. https://doi.org/10.1016/j.bmcl.2009.03.043

Cheung J, Rudolph MJ, Burshteyn F, Cassidy MS, Gary EN, Love J, . . . Height JJ (2012) Structures of human acetylcholinesterase in complex with pharmacologically important ligands. J Med Chem 55(22):10282–10286. https://doi.org/10.1021/jm300871x

Choi D-Y, Lee Y-J, Hong JT, Lee H-J (2012) Antioxidant properties of natural polyphenols and their therapeutic potentials for Alzheimer's disease. Brain Res Bull 87(2–3):144–153. https://doi.org/10.1016/j.brainresbull.2011.11.014

Cumming JN, Smith EM, Wang L, Misiaszek J, Durkin J, Pan J, . . . Stamford AW (2012) Structure based design of iminohydantoin BACE1 inhibitors: identification of an orally available, centrally active BACE1 inhibitor. Bioorg Med Chem Lett 22(7):2444–2449. https://doi.org/10.1016/j.bmcl.2012.02.013

Drwal MN, Banerjee P, Dunkel M, Wettig MR, Preissner R (2014) ProTox: a web server for the in silico prediction of rodent oral toxicity. Nucleic Acids Res 42(W1):W53–W58. https://doi.org/10.1093/nar/gku401

Fedorov R, Hartmann E, Ghosh DK, Schlichting I (2003) Structural basis for the specificity of the nitric-oxide synthase inhibitors W1400 and N -propyl-L-Arg for the inducible and neuronal isoforms. J Biol Chem 278(46):45818–45825. https://doi.org/10.1074/jbc.M306030200

Feng Y, Wang X (2012) Antioxidant therapies for Alzheimer's disease. Oxidative Med Cell Longev 2012:1–17. https://doi.org/10.1155/2012/472932

Furukawa H, Gouaux E (2003) Mechanisms of activation, inhibition and specificity: crystal structures of the NMDA receptor NR1 ligand-binding core. EMBO J 22(12):2873–2885. https://doi.org/10.1093/emboj/cdg303

Grundman M, Grundman M, Delaney P (2002) Antioxidant strategies for Alzheimer's disease. Proc Nutr Soc 61(2):191–202. https://doi.org/10.1079/PNS2002146

Johansson MU, Zoete V, Michielin O, Guex N (2012) Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. BMC Bioinforma 13(1):173. https://doi.org/10.1186/1471-2105-13-173

Klugman A, Naughton DP, Isaac M, Shah I, Petroczi A, Tabet N (2012) Antioxidant enzymatic activities in Alzheimer's disease: the relationship to acetylcholinesterase inhibitors. J Alzheimers Dis JAD 30(3):467–474. https://doi.org/10.3233/JAD-2012-120124

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46(1–3):3–26

Manivannan J, Silambarasan T, Kadarkarairaj R, Raja B (2015) Systems pharmacology and molecular docking strategies prioritize natural molecules as cardioprotective agents. RSC Adv 5(94):77042–77055. https://doi.org/10.1039/C5RA10761J

Meng X-Y, Zhang H-X, Mezei M, Cui M (2011) Molecular docking: a powerful approach for structure-based drug discovery. Curr Comp Aided Drug Des 7(2):146–157. https://doi.org/10.2174/157340911795677602

Montine TJ, Amarnath V, Martin ME, Strittmatter WJ, Graham DG (1996) E-4-hydroxy-2-nonenal is cytotoxic and cross-links cytoskeletal proteins in P19 neuroglial cultures. Am J Pathol 148 (1):89–93

Moroy G, Martiny VY, Vayer P, Villoutreix BO, Miteva MA (2012) Toward in silico structure-based ADMET prediction in drug discovery. Drug Discov Today 17(1–2):44–55. https://doi.org/10.1016/j.drudis.2011.10.023

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 30(16):2785–2791. https://doi.org/10.1002/jcc.21256

Newman DJ, Cragg GM (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. J Nat Prod 75(3):311–335. https://doi.org/10.1021/np200906s

Niu X, Umland S, Ingram R, Beyer BM, Liu Y-H, Sun J, . . . Orth P (2006) IK682, a tight binding inhibitor of TACE. Arch Biochem Biophys 451(1):43–50. https://doi.org/10.1016/j.abb.2006.03.034

Sayre LM, Zelasko DA, Harris PL, Perry G, Salomon RG, Smith MA (1997) 4-Hydroxynonenal-derived advanced lipid peroxidation end products are increased in Alzheimer's disease. J Neurochem 68(5):2092–2097

Selinsky BS, Gupta K, Sharkey CT, Loll PJ (2001) Structural analysis of NSAID binding by prostaglandin H2 synthase: time-dependent and time-independent inhibitors elicit identical enzyme conformations. Biochemistry 40(17):5172–5180

Sharma V, Sharma PC, Kumar V (2016) In silico molecular docking analysis of natural pyridoacridines as anticancer agents. Adv Chem 2016:1–9. https://doi.org/10.1155/2016/5409387

Smith CD, Carney JM, Starke-Reed PE, Oliver CN, Stadtman ER, Floyd RA, Markesbery WR (1991) Excess brain protein oxidation and enzyme dysfunction in normal aging and in Alzheimer disease. Proc Natl Acad Sci U S A 88(23):10540–10543

Smith MA, Taneda S, Richey PL, Miyata S, Yan SD, Stern D, . . . Perry G (1994) Advanced maillard reaction end products are associated with Alzheimer disease pathology. Proc Natl Acad Sci USA 91(12):5710–5714

Smith MA, Rudnicka-Nawrot M, Richey PL, Praprotnik D, Mulvihill P, Miller CA, . . . Perry G (1995) Carbonyl-related posttranslational modification of neurofilament protein in the neurofibrillary pathology of Alzheimer's disease. J Neurochem 64(6):2660–2666

Smith MA, Richey Harris PL, Sayre LM, Beckman JS, Perry G (1997) Widespread peroxynitrite-mediated damage in Alzheimer's disease. J Neurosci Off J Soc Neurosci 17(8):2653–2657

Sung B-J, Hwang KY, Jeon YH, Lee JI, Heo Y-S, Kim JH, . . . Cho JM (2003) Structure of the catalytic domain of human phosphodiesterase 5 with bound drug molecules. Nature 425(6953):98–102. https://doi.org/10.1038/nature01914

Umukoro S, Adewole FA, Eduviere AT, Aderibigbe AO, Onwuchekwa C (2014) Free radical scavenging effect of donepezil as the possible contribution to its memory enhancing activity in mice. Drug Res 64(5):236–239. https://doi.org/10.1055/s-0033-1357126

Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD (2002) Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 45(12):2615–2623

Vecchio AJ, Malkowski MG (2011) The structure of NS-398 bound to cyclooxygenase-2. J Struct Biol 176(2):254–258. https://doi.org/10.1016/j.jsb.2011.07.019

Wandhammer M, de Koning M, van Grol M, Loiodice M, Saurel L, Noort D, . . . Nachon F (2013) A step toward the reactivation of aged cholinesterases – crystal structure of ligands binding to aged human butyrylcholinesterase. Chem Biol Interact 203(1):19–23. https://doi.org/10.1016/j.cbi.2012.08.005

Watermeyer JM, Kröger WL, O'Neill HG, Sewell BT, Sturrock ED (2008) Probing the basis of domain-dependent inhibition using novel ketone inhibitors of angiotensin-converting enzyme. Biochemistry 47(22):5942–5950. https://doi.org/10.1021/bi8002605

Wolohan PRN, Clark RD (2003) Predicting drug pharmacokinetic properties using molecular interaction fields and SIMCA. J Comput Aided Mol Des 17(1):65–76

Zhao YH, Abraham MH, Le J, Hersey A, Luscombe CN, Beck G, . . . Cooper I (2002) Rate-limited steps of human oral absorption and QSAR studies. Pharm Res 19(10):1446–1457

Chapter 14
In Silico Biochemical Pathways for Bacterial Metabolite Synthesis

Murtaza Ali, Shahnawaz Ali, and Romana Ishrat

Abstract The search of an alternative is always a major concern of human for sustaining life effectively. The green flora described by the agriculture is the source of many such life attaining processes and products that are essential for human population. In addition to this, the constantly growing number of *Homo sapiens* has to be fed with increase yield of agriculture products. To meet the demands of growing population and relieve the pressure of yield, the use of fertilizers comes into action, while the constant use of chemical fertilizers has deteriorated the heath of soil, environment, and human collectively calling "phytobiome." Thus, the urge of finding alternatives to replace the toxic chemical fertilizers has given a way to search exhaustively the naturally occurring microbiomes for their beneficial effect on the agri-flora. Moreover, the available advancements in the computational and system-level approaches with omics data have provided us the genomes and also genome-level metabolic models for many beneficial/effective bacteria. The naturally synthesized metabolites (primary and secondary) can be easily exploited nowadays for any intended use in the fields as inoculants or bio fertilizers. In addition the available kinetic model has paved the way to commercially synthesize desired metabolite (through amendments in pathway either genetic or environmental) on large scale as biofuels, etc. Despite of these advances, several challenges still coexist with approaches that have to be exploited in the near future. Some of the challenges have been discussed in present work with a brief account of in silico kinetic models available.

Author contributed equally with all other contributors. Murata Ali and Shahnawaz Ali

M. Ali
Department of Stem Cell and Cell Culture, Center for Advance Research, King George Medical University, Lucknow, India

S. Ali (✉) · R. Ishrat (✉)
Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi, India
e-mail: shahnawaz127339@st.jmi.ac.in; rishrat@jmi.ac.in

D. K. Choudhary et al. (eds.), *In Silico Approach for Sustainable Agriculture*,
https://doi.org/10.1007/978-981-13-0347-0_14

14.1 Introduction

Major interest from both academic and industrial point of views, nowadays, lies in obtaining insight in going "green." Today, the search of alternatives has pounded a way to research on the so-called green solvents that are bio-based and nontoxic and have good environmental footprint. The use of synthetic chemicals (like fertilizers, pesticides, and so on) has provided us good yield in respect to cost and quantity but had also paved the way of toxic chemicals to major concern (as, for instance, biomagnification). So the need of so-called sustainable and nontoxic agriculture has to be emphasized by knowing the underlying metabolisms of naturally available alternatives that are dynamic and in some sense nonrenewable sources. Thus, the idea of sustained agriculture goes parallel with the availability of naturally occurring feedstock (like agrobacterium) and in the underlying mechanism of their metabolite synthesis.

In this era of omics technologies, one can find a burst of information on many model and non-model organisms. The availability of massively parallel monitoring of cellular components on the transcriptional (*transcriptomics*) and translational (*proteomics*) level now allows for a quantification of metabolic compounds (*metabolomics*) on a large scale. The development of novel methods and computational methods has facilitated the analysis of the intertwined processes taking place in living cells. These new and quantitative forms of analysis can be attributed to the emergent field of systems biology and are based on data generated from high-throughput techniques and computational tools; see Tables 14.2 and 14.3.

Effective microbes can be utilized in agriculture for various purposes, as, for instance, using inoculants for biological nitrogen fixation to improve crop quality and yields. The fundamental consideration in applying beneficial microbes is to enhance their synergistic effect (Parr et al. 1994; Higa and Wididana 1991a). Application of these microbes also depends on the type and nature of the environment they are applied in thus ensuring great diversity. For example, combinations of various crop residues, animal manures, green manures, and municipal wastes will provide a higher level of microbial diversity. The reason for this is that each of these organic materials has its own unique indigenous microflora which can greatly affect the resident soil microflora after they are applied, at least for a limited period.

This chapter discusses in brief the available pathways (both metabolic and non-metabolic) that help in the synthesis of some major metabolite in bacterial population responsible for flourishing *ager*-flora in a natural way. These include class of metabolites that are nowadays synthesized as a part of going green and produced commercially to oust the use of toxic fertilizers from use. This chapter will discuss in specific the biochemical pathways that are used by bacteria for metabolite synthesis with respect to agriculture. In addition to this, it also gives a brief account of their recent application and research on the go, with an account of some unknown or ubiquitous questions in the research community.

14.2 Connection Between Plants and Microbes: General View

The two terms that are most promptly used in describing these connections (either beneficial or pathogenic) are "rhizosphere" and "phytobiomes."

Rhizosphere by definition means the edge involving the interaction of root systems and surrounding microbes. It is estimated to be the most complex habitat on earth involving an approximate of 10E11 microbial cells (e.g., archaea, bacteria, eukaryotes, and viruses). Rhizosphere microbiome in specific sense represents plant-specific (in particular) community of rhizosphere microbes that are fueled by plant exudates (Doornbos et al. 2012). Since, the plants affect surrounding soil by the composition of their exudates, which in return control the growth and fitness by their structural and functional properties (Burns 2010; Morgan and Whipps 2000).

Phytobiome is a broad term that refers to the collection of plant itself, the environment, and all micro- and macroorganisms living in, on, or around the plant (including rhizospheres) that influence plant growth, development, and evolutionary dynamics. In 2014 the American Phytopathological Society (APS) has decided to take a leap forward and considered the strategies for shaping a new paradigm for crop improvement focusing on phytobiome-based approaches. This approach may instrumentalize the system-level understanding of the diverse interacting components in context to sustainable agriculture spanning multiple disciplines, species, and environments. This omics data produced at very fine scales of space and time will be coupled across to provide novel insight into the interaction networks that facilitate plant productivity (Phytobiomes: A Roadmap for Research and Translation 2016).

14.3 Beneficial Bacteria (Microflora) of Agricultural Importance

The uniqueness of any microorganisms lies in their biosynthetic capabilities under a given set of environmental conditions. Thus, holding these capabilities had made these *micro* creatures preferred candidates to solve various challenging problems in many fields. Over the past century, these microorganisms have been used in different ways for the advances seen in food safety and quality, medical technology, and most importantly agriculture. Many of these recorded achievements have been possible due to the in-depth understanding of metabolism and straightforward application of chemical and physical engineering methods with practical and economic feasibility.

In the recent years, it is found that successfully using microbial technologies and constantly producing their beneficial effects are not possible. The fact here is microorganisms are only operative once they are offered appropriate and optimum conditions for metabolizing their substrates that include water (available), oxygen (depending on whether the microorganisms are obligate aerobes or facultative anaerobes), pH, and temperature of their environment. While the availability of

inoculants has increased as microorganisms are useful in eliminating problems associated with the use of chemical fertilizers and pesticides, they are widely used in farming and organic agriculture (Higa 1991; Parr et al. 1994).

For many years, the scientific community (soil microbiologists and microbial ecologists) has tended to segregate microorganisms as beneficial or harmful according to their functions and how they affect soil quality, plant growth and yield, and plant health (see Box 14.1). For instance, microorganisms that can fix atmospheric nitrogen, decompose organic wastes and residues, detoxify pesticides, suppress plant diseases and soil-borne pathogens, enhance nutrient cycling, and produce bioactive compounds such as vitamins, hormones, and enzymes that stimulate plant growth are termed as "beneficial." Those that can induce plant diseases, stimulate soil-borne pathogens, immobilize nutrients, and produce toxic and putrescent substances effecting plant growth and health are called "harmful." The more effective grouping of beneficial and harmful organism is given by Higa (Higa and Wididana 1991a, b) referring them as EM (or effective microorganisms) (Higa 1994). EM cultures contain selected species of microorganisms (that may include lactic acid bacteria and yeasts and smaller numbers of photosynthetic bacteria, actinomycetes, and other types of organisms) to improve soil quality, its health, and the growth, yield, and quality of crops. So the term "beneficial microorganisms" is a general way to designate a large group of often unknown or ill-defined microorganisms that interact favorably in soils and with plants to render beneficial effects which are sometimes difficult to predict. We use the term "effective microorganisms" or EM to denote specific mixed cultures of known, beneficial microorganisms that are being used effectively as microbial inoculants (Higa and Wididana 1991a, b; Higa 1994).

Box 14.1: Some Common Functions of Beneficial and Harmful Microorganisms (Higa and Wididana 1991a)

Beneficial	Harmful
Fixation of atmospheric nitrogen	Initiation of plant diseases
Decomposition of organic wastes and residues	Stimulating soil-borne pathogens
Suppression of soil-borne pathogens	Immobilization of plant nutrients
Recycling and increased availability of plant nutrients	Inhibition of seed germination
Degradation of toxicants including pesticides	Inhibition of plant growth and development
Production of antibiotics and other bioactive compounds	Production of phytotoxic substances
Production of simple organic molecules for plant uptake	
Complexation of heavy metals to limit plant uptake	
Solubilization of insoluble nutrient sources	

A class of bacteria called plant growth-promoting rhizobacteria (PGPR) forms the potential group to aid in controlling plant pathogens and helping in to improve soil fertility and crop yield for beneficial agricultural. Several compounds like phenazines, pyoluteorin, and pyrrolnitrin as well as some enzymes, metabolites, antibiotics, etc. produced by PGPR are vital in colonizing rhizosphere. PGPR may overcome a variety of deleterious effects (see Box 14.2 for some) (Babalola 2010).

Box 14.2: PGPR Compensating Effects

Reduction in plant growth caused by weed infestation (Babalola et al. 2007)
Drought stress (Zahir et al. 2008)
Heavy metals (Kumar et al. 2009)
Salt stress (Egamberdieva 2008; Kaymak et al. 2009)
Some other unfavorable environmental conditions

Some of the prominent PGPR as given by Babalola in 2010 article (Babalola 2010) belonging to these genera include *Acetobacter*, *Acinetobacter*, *Alcaligenes*, *Arthrobacter*, *Azoarcus*, *Azospirillum*, *Azotobacter*, *Bacillus*, *Beijerinckia*, *Burkholderia*, *Derxia*, *Enterobacter*, *Gluconacetobacter*, *Herbaspirillum*, *Klebsiella*, *Ochrobactrum*, *Pantoea*, *Pseudomonas*, *Rhodococcus*, *Serratia*, *Stenotrophomonas*, and *Zoogloea* and have been subject of extensive research throughout the years. For more specific effects of PGPR to a particular plant (see Table 1 of Babalola (2010).

14.4 Production of Metabolites in Bacteria

Metabolites in agricultural sense are the "end" or "intermediate" and beneficial products of cellular regulatory mechanisms. Bacterial metabolites, for instance, show a variety of biological properties with most of them being pharmacologically important. These metabolites assemble as a result of cascade of enzymatic activity using various sources like organic, inorganic, and de novo. These synthesized metabolites can be classified broadly into primary and secondary. Primary metabolites include those compounds (ethanol, lactic acid, and certain amino acids) that are directly involved in intrinsic functions (like normal growth, development, and reproduction), while that of secondary metabolites has a relational function. Some common secondary metabolites include alkaloids, antibiotics, nucleosides, phenazines, quinolines, terpenoids, peptides, and growth factors.

Secondary metabolites are more focused in a research as they are involved in direct relation of the crop production and development. These are assembled using pathways involving chiral, nonproteinogenic amino acids and using acetate units from acyl-CoA. These naive metabolites are further modified for intended functions. Late phase of bacterial growth witnesses several biochemical pathways for the

production of secondary metabolites. They have many direct and indirect effects on bacteria and its associates. Here we will highlight some of these pathways involved using nonribosomal peptide synthase, polyketide synthase, shikimate pathways, and β-lactam pathways.

14.5 In Silico Models of Metabolic Models: Regulation and Stability

Despite of the overpowering complexity, living organisms are chemical systems run by a set of chemical reactions taking place in an aqueous environment. This set of chemical reactions raises behaviors that enable this chemical system (say cells) and organisms to grow and to reproduce (Alberts et al. 2002). Decoding the design of this complex underlying mechanism of physicochemical processes remains the greatest challenges of the time.

Since the start of the 1990s, the availability of high-throughput technologies has enabled us to model the integrated organization and functioning of metabolic processes; see Table 14.1. Metabolism being a complex process can be broadly subdivided into two opposing streams of interconversions: *catabolism*, the breakdown of macromolecules into smaller units, and *anabolism*, the production of new macromolecules from units obtained through catabolism. The organization of metabolism can be described as a *bowtie* architecture (Ma and Zeng 2003; Csete and Doyle

Table 14.1 Selected examples of kinetic models of metabolism

Pathway	Author/ref	Year
Horseradish peroxidase	Chance (1943)	1943
Carbohydrate metabolism	Wright et al. (1968)	1968
Glycolysis (energy)	Selkov (1968)	1968
Glycolysis and fermentation	Galazzo and Bailey (1990)	1990
TCA cycle	Wright et al. (1992)	1992
TCA cycle	ElMansi et al. (1994)	1994
Carbohydrate metabolism	Wright and Albe (1994)	1994
Central metabolism	Rizzi et al. (1997)	1997
Yeast glycolysis	Teusink et al. (2000)	1999
Pentose phosphate pathway	Vaseghi et al. (1999)	1999
Glycolysis	Wolf et al. (2000)	2000
Threonine pathway	Chassagnole et al. (2001)	2001
Glycolysis	Hynne et al. (2001)	2001
Central carbon metabolism	Chassagnole et al. (2002)	2002
Glycolysis (*L. lactis*)	Hoefnagel et al. (2002)	2002
Yeast glycolysis	Klipp et al. (2005)	2005

Note that the list is not comprehensive. For more examples see also the model repositories listed in Table 14.2

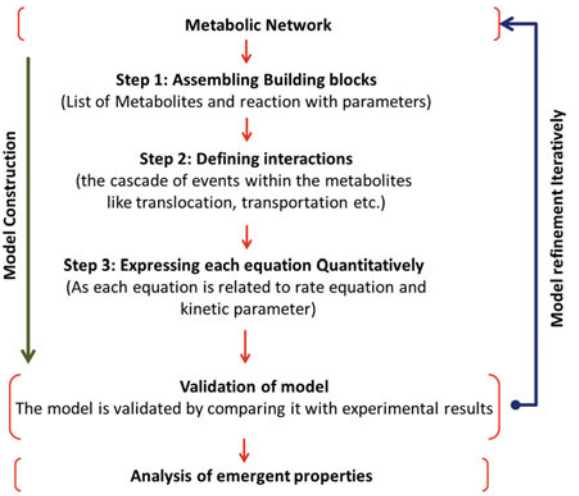


Fig. 14.1 An overview of scheme followed to construct the mathematical metabolic model. However, in reality many approximations are made to minimize the divergence from reality (Wiechert and Takors 2004)

2004); see Fig. 14.1 for an overview. In order to understand the organization of organism, mathematical modeling with formal analysis can serve a pragmatic role. Thus our aim of purposefully altering metabolic processes for desired effect on agriculture can be elucidated by constructing large-scale metabolic models some of them shown in Table 14.1, while many of them can be found in the repositories given in Table 14.2 and modeled using tools in Table 14.3. This will give access to overcome a number of challenges of outstanding relevance, ranging from global crop supply to the synthesis of biofuels, and directly relates to our ability to utilize microbial or plant metabolic pathways in a purposeful way.

14.6 Mechanistic Kinetic Models of Metabolic Pathways

The straightforward and well-known approach to model metabolic pathways is to represent metabolic processes in terms of ordinary differential equations (ODEs). The set chemical reactions defining processes that capture changes in metabolite concentrations are described by a mass balance equation incorporating kinetic details of reaction mechanisms and their associated kinetic parameters. Despite recent attempts to construct "genome-scale" kinetic models of cellular metabolism (Tomita et al. 1999; Ishii et al. 2004; Jamshidi and Palsson 2008; Slepchenko and Schaff 2003), the kinetic modeling is currently often limited to smaller or individual pathways. The kinetic models involve several hundreds of equations. The quest for genome-scale dynamic models has given a paradigm toward comprehensive, predictive, and "in silico" models (Blazeck and Alper 2010). In addition to it, minimal models have been a trademark of theoretical biology during the past decades. These minimal models have been used extensively, to understand oscillatory phenomena of glycolysis in various microbes.

Table 14.2 Useful resources for information on kinetic models, including pathways present in databases and model repositories

Names	Description	Link
KEGG	Kyoto encyclopedia of genes and genomes	http://www.genome.jp/kegg/
METACYC	Database of metabolic pathways and enzymes	http://metacyc.org/
Reactome	A curated resource of core pathways in human biology	http://www.reactome.org/
BRENDA	The BRaunschweig ENzyme DAtabase	http://www.brendaenzymes.info/
SABIO RK	System for the analysis of biochemical pathways reaction kinetics	http://sabio.villabosch.de/SABIORK/
NIST	Thermodynamics of enzyme catalyzed reactions	http://xpdb.nist.gov/enzyme thermodynamics/
BioModels	A database of annotated published models	http://www.ebi.ac.uk/biomodels/
JWS online	A tool for simulation of kinetic models from a curated database	http://www.jjj.bio.vu.nl/
CellML	XML-based language to store and exchange models	http://www.cellml.org/
CellML repository	Repository	http://www.cellml.org/models
SBML	The systems biology markup language	http://www.sbml.org/
MIRIAM	Standardized minimal information in the annotation of models	http://www.ebi.ac.uk/compneursrv/miriam/
Systems biology	A portal site for systems biology	http://systemsbiology.org
SMBP	Secondary metabolite bioinformatics portal	http://www.secondarymetabolites.org

Table 14.3 Some of the software packages for kinetic simulations

Name	URL	References
Copasi	http://www.copasi.org/	Hoops et al. (2006)
CellDesigner	http://www.celldesigner.org/	Funahashi et al. (2008)
ECELL	http://www.ecell.org/	Tomita et al. (1999)
Cellware	http://www.cellware.org/	Dhar et al. (2004)
JDesigner/Jarnac	http://sbw.kgi.edu/sbwWiki/sysbio/jdesigner/	Sauro et al. (2003)
SB toolbox	http://www.sbtoolbox2.org/	Schmidt and Jirstrand (2006)
SBRT	http://www.bioc.uzh.ch/wagner/software/SBRT/	Wright and Wagner (2008)
SEQL-NRPS	http://services.birc.au.dk/seql-nrps/	Knudsen et al. (2015)

A more comprehensive list can be found (see Weber and Kim 2016) and for more SMBP (http://www.secondarymetabolites.org)

The time-dependent behavior of a metabolic network consists of m metabolic reactants (metabolites) interacting via a set of n biochemical reactions. Each metabolite S_i is the concentration following $S_i(t) \geq 0$, usually measured in moles/volume. Each such reaction is guided by two quantities: (i) the stoichiometric coefficients N_{ij} and (ii) a (often nonlinear) function $\nu(S, k)$ that specifies the rate of the reaction as a function of the concentration vector S and a set of kinetic parameters k.

Once the stoichiometric coefficients N_{ij} and rate functions ν_j for all reactions is obtained, the time-dependent concentration of a metabolic reactant $S_i(t)$ is described by the dynamic mass balance equation in matrix notation as follows:

$$\frac{dS(t)}{dt} = N\nu(S, k) \tag{14.1}$$

The above equation describes the variation of a metabolite concentration over time. A stationary and time-invariant state of metabolite concentrations S^0 (steady state) is regarded as by the steady-state condition.

$$\frac{dS(t)}{dt} = N\nu(S, k) = 0 \tag{14.2}$$

Usually the steady-state condition describes the stationary nonequilibrium state, with nonzero net flux and positive entropy production. The stoichiometric matrix used to solve kinetic models is an *invariant* property of metabolic reaction networks, that is, it doesn't consider physiological conditions like pH, temperature, etc. For a schematic workflow, see Fig. 14.1. For more exhaustive information on modeling and analysis techniques, see detailed description of Steuer and Junkeret (2010) and Ao et al. (2008); it has comprehensively discussed stoichiometric and topological analysis of metabolic models with limitation.

14.7 Future Perspectives and Conclusions

Over the past decade, an increase in computational modeling has provided an important contribution to the understanding of metabolism of the cell in all organisms. The field of systems biology, mathematical modeling, and other computational approaches had overtaken the central stage of molecular biology. A system-level understanding of bacterial metabolism has become critical for assessing its potentials and weaknesses in order to exploit them for desired application. However, global analysis of metabolism is not an easy task. The plethora of reactions that governs metabolism of any microbe depends on a variety of other factors also. Even the known reactions to us and enzymes can interact in many different ways to give rise to different tasks. For instance, PGPR that are free-living bacteria have more than one mechanism to effect effectively plant growth, production of enzymes, bioactive factors, antibiotics, and metabolites.

Thus one can attempt to conduct system-level studies mainly because of the availability of various forms of omics data. New algorithms are developed for integration of new biological data that in turn can be used to mathematically model, analyze, and predict bacterial behavior with increasing accuracy. Eventually the inclusion of detailed modeling of transcription and translation processes could expand the scope of system-level models beyond metabolism. These developments could in the near future make in silico system-level analysis a transformative technology.

In spite of the significant advances on computational approaches, there are still several challenges exist that have to be taken into consideration in the near future. Even for the very well-studied secondary metabolite classes of NRP (non-ribosomally synthesized peptide) pathways, prediction is incomplete because the appropriate biochemical data is not adequately available. In particular, for machine learning-based approaches, the availability of large-scale biochemical data requires good training models which are very limiting. Another more important unsolved problem is an inaccurate prediction of gene cluster.

Analyses that involve integration of diverse data forms (e.g., different omics data) suffer from an ill integration of different tools with some specific requirement of input and output formats; all these make hurdles and hamper the usage and analysis difficult for researchers. In fact, this is a chronic problem in bioinformatics and systems biology. Genome mining and other omics-based approaches have become essential technologies accompanying the classical approaches of detection. This famous trend now has displayed with increasing number of new and improved chemi- and bioinformatic tools and databases, which bridges gap between in silico and in vitro/vivo work.

Acknowledgments We wish to thank all the authors for helpful discussions and for preparation of figures. SA is supported by ICMR-SRF.

References

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002) Molecular biology of the cell, 4th edn. Garland Science (Taylor & Francis Group), New York

Ao P, Lee LW, Lidstrom M, Yin L, Zhu XM (2008) Towards kinetic modeling of global metabolic networks with incomplete experimental input on kinetic parameters. arXiv preprint arXiv:0808.0220

Babalola OO (2010) Beneficial bacteria of agricultural importance. Biotechnol Lett 32:1559–1570

Babalola OO, Sanni AI, Odhiambo GD, Torto B (2007) Plant growth-promoting rhizobacteria do not pose any deleterious effect on cowpea and detectable amounts of ethylene are produced. World J Microbiol Biotechnol 23(6):747–752

Blazeck J, Alper H (2010) Systems metabolic engineering: genome-scale models and beyond. Biotechnol J 5(7):647–659

Burns RG (2010) Albert Rovira and a half-century of rhizosphere research. In: Proceedings of the Rovira rhizosphere symposium, p 1

Chance B (1943) The kinetics of the enzyme substrate compound of peroxidase. J Biol Chem 151 (2):553–577
Chassagnole C, Raïs B, Quentin E, Fell DA, Mazat JP (2001) An integrated study of threonine pathway enzyme kinetics in *escherichia coli*. Biochem J 356(Pt 2):415–423
Chassagnole C, Noisommit Rizzi N, Schmid JW, Mauch K, Reuss M (2002) Dynamic modeling of the central carbon metabolism of *Escherichia coli*. Biotechnol Bioeng 79(1):53–73
Csete M, Doyle J (2004) Bow ties, metabolism and disease. Trends Biotechnol 22(9):446–450
Dhar P, Meng TC, Somani S, Ye L, Sairam A, Chitre M, Hao Z, Sakharkar K (2004) Cellware-a multi-algorithmic software for computational systems biology. Bioinformatics 20 (8):1319–1321
Doornbos RF, van Loon LC, Bakker PA (2012) Impact of root exudates and plant defense signaling on bacterial communities in the rhizosphere. A Rev Agron Sustain Dev 32(1):227–243
Egamberdieva D (2008) Plant growth promoting properties of rhizobacteria isolated from wheat and pea grown in loamy sand soil. Turk J Biol 32(1):9–15
ElMansi EMT, Dawson GC, Bryce CFA (1994) Steady-state modelling of metabolic flux between the tricarboxylic cycle and the glyoxylate bypass in *Escherichia coli*. CABIOS (now Bioinformatics) 10(3):295–299
Funahashi A, Matsuoka Y, Jouraku A, Morohashi M, Kikuchi N, Kitano H (2008) Cell designer 3.5: a versatile modeling tool for biochemical networks. Proc IEEE 96(8):1254–1265
Galazzo JL, Bailey JE (1990) Fermentation pathway kinetics and metabolic flux control in suspended and immobilized *saccharomyces cerevisiae*. Enzym Microb Technol 12(3):162–172
Higa T (1991) Effective microorganisms: a biotechnology for mankind. In: Parr JF, Hornick SB, Whitman CE (eds) Proceedings of the first international conference on Kyusei Nature Farming. U.S. Department of Agriculture, Washington, DC, pp 8–14
Higa T (1994) Effective microorganisms: a new dimension for nature farming. In: Parr JR, Hornick SB, Simpson ME (eds) Proceedings of the second international conference on Kyusei Nature Farming. U.S. Department of Agriculture, Washington, DC, pp 20–22
Higa T, Wididana GA (1991a) The concept and theories of effective microorganisms. In: Parr, Hornick SB, Whitman CE (eds) Proceedings of the first international conference on Kyusei Nature Farming. U.S. Department of Agriculture, Washington, DC, pp 118–124
Higa T, Wididana GA (1991b) Changes in the soil microflora induced by effective microorganisms. In: Parr JF, Hornick SB, Whitman CE (eds) Proceedings of the first international conference on Kyusei Nature Farming. U.S. Department of Agriculture, Washington, DC, pp 153–162
Hoefnagel MHN et al (2002) Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis. Microbiology 148 (Pt 4):1003–1013
Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, Singhal M, Xu L, Mendes P, Kummer U (2006) COPASI-a complex pathway simulator. Bioinformatics 22(24):3067–3074
Hynne F, Danø S, Sørensen PG (2001) Fullscale model of glycolysis in *Saccharomyces cerevisiae*. Biophys Chem 94:121–163
Ishii N, Robert M, Nakayama Y, Kanai A, Tomita M (2004) Toward large scale modeling of the microbial cell for computer simulation. J Biotechnol 113:281–294
Jamshidi N, Palsson BO (2008) Formulating genome scale kinetic models in the postgenome era. Mol Syst Biol 4:171
Kaymak HC, Guvenc I, Yarali F, Donmez MF (2009) The effects of bio-priming with PGPR on germination of radish (*Raphanus sativus* L.) seeds under saline conditions. Turk J Agric For 33 (2):173–179
Klipp E et al (2005) Integrative model of the response of yeast to osmotic shock. Nat Biotechnol 23 (8):975–982
Knudsen M, Sondergaard D, Tofting-Olesen C, Hansen FT, Brodersen DE, Pedersen CN (2015) Computational discovery of specificity-conferring sites in non-ribosomal peptide synthetases. Bioinformatics. https://doi.org/10.1093/bioinformatics/btv600

Kumar KV, Srivastava S, Singh N, Behl HM (2009) Role of metal resistant plant growth promoting bacteria in ameliorating fly ash to the growth of *Brassica juncea*. J Hazard Mater 170(1):51–57
Ma H, Zeng AP (2003) Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics 19(2):270–277
Morgan JAW, Whipps JM (2000) Methodological approaches to the study of rhizosphere carbon flow and microbial population dynamics. In: The rhizosphere: biochemistry and organic substance at the soil-plant interface: biochemistry and organic substance at the soil-plant interface, p 373
Parr JF, Hornick SB, Kaufman DD (1994) Use of microbial inoculants and organic fertilizers in agricultural production. In: Proceedings of the international seminar on the use of microbial and organic fertilizers in agricultural production. Published by the Food and Fertilizer Technology Center, Taipei, Taiwan
Phytobiomes: A Roadmap for Research and Translation (2016) American Phytopathological Society, St. Paul. www.phytobiomes.org/roadmap
Rizzi M et al (1997) In vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae*: Ii. Mathematical model. Biotechnol Bioeng 55:592–608
Sauro HM, Hucka M, Finney A, Wellock C, Bolouri H, Doyle J, Kitano H (2003) Next generation simulation tools: the systems biology workbench and BioSPICE integration. Omics A J Integr Biol 7(4):355–372
Schallau K, Junker BH (2010) Simulating plant metabolic pathways with enzyme-kinetic models. Plant Physiol 152(4):1763–1771
Schmidt H, Jirstrand M (2006) Systems biology toolbox for matlab: a computational platform for research in systems biology. Bioinformatics 22(4):514–515
Selkov EE (1968) Self oscillations in glycolysis. Eur J Biochem 4:79–86
Slepchenko BM, Schaff JC, Macara I, Loew LM (2003) Quantitative cell biology with the virtual cell. Trends Cell Biol 13(11):570–577
Teusink B et al (2000) Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. Eur J Biochem 267(17):5313–5329
Tomita M, Hashimoto K, Takahashi K, Shimizu TS, Matsuzaki Y, Miyoshi F, Saito K, Tanida S, Yugi K, Venter JC, Hutchison CA 3rd (1999) E-CELL: software environment for whole-cell simulation. Bioinformatics (Oxford, England) 15(1):72–84
Vaseghi S, Baumeister A, Rizzi M, Reuss M (1999) In vivo dynamics of the pentose phosphate pathway in *Saccharomyces cerevisiae*. Metab Eng 1(2):128–140
Weber T, Kim HU (2016) The secondary metabolite bioinformatics portal: computational tools to facilitate synthetic biology of secondary metabolite production. Synth Syst Biotechnol 1:69–79
Wiechert W, Takors R (2004) Validation of metabolic models: concepts, tools, and problems. In: Kholodenko BN, Westerhoff HV (eds) Metabolic engineering in the post genomic era. Horizon Bioscience, UK, pp 277–320
Wolf J, Passarge J, Somsen OJG, Snoep JL, Heinrich R, Westerhoff HV (2000) Transduction of intracellular and intercellular dynamics in yeast glycolytic oscillations. Biophys J 78:1145–1153
Wright BE, Albe KR (1994) Carbohydrate metabolism in *Dictyostelium discoideum*: I. model construction. J Theor Biol 169(3):231–241
Wright J, Wagner A (2008) The systems biology research tool: evolvable open source software. BMC Syst Biol 2:55
Wright B, Simon W, Walsh T (1968) A kinetic model of metabolism essential to differentiation in *Dictyostelium discoideum*. Proc Natl Acad Sci 60:644–651
Wright BE, Butler MH, Albe KR (1992) Systems analysis of the tricarboxylic acid cycle in dictyostelium discoideum. i. The basis for model construction. J Biol Chem 267(5):3101–3105
Zahir ZA, Munir A, Asghar HN, Shaharoona B, Arshad M (2008) Effectiveness of rhizobacteria containing ACC deaminase for growth promotion of peas (*Pisum sativum*) under drought conditions. J Microbiol Biotechnol 18(5):958–963

Chapter 15
In Silico Characterization of Plant Secondary Metabolites

A. Sabitha Rani, G. Neelima, Rupasree Mukhopadhyay, K. S. N. Jyothi, and G. Sulakshana

Abstract Plants are a rich source of chemical compounds which serve as food, colors, fragrances', flavors, medicines, etc. Plant secondary metabolites are widely used in food technology, industry, and medicinal preparations and play a vital role in plant-environment interactions. These metabolites have unique characteristics which make them as important candidates for discovery of new drugs and "lead" molecules. So far the major lacuna in the area of plant metabolite research is the identification and characterization of the secondary metabolites and their biosynthetic mechanisms. With an upsurge in the demand for plant metabolites, the advanced "omics technologies" are most sought after for a faster research and better characterization of the natural products. With the advent of the advanced bioinformatics, genomics, and proteomics and the synergy between combinatorial chemistry and structure-based drug design, the process of characterizing secondary metabolites as lead molecules for drug design has been revolutionized. The scientific community is now witnessing a newer, faster, and sophisticated approach to drug discovery with the aid of in silico characterization methods. This chapter, thus, focuses on the general steps to be followed in the *in silico* characterization of plant secondary metabolites, starting from literature mining, virtual screening, structural characterization, ADMET screening, and structure-based drug designing.

Keywords Secondary metabolites · Virtual screening · Combinatorial chemistry · *In silico* characterization · Ligand-based screening · Docking · Drug designing

A. Sabitha Rani (✉) · G. Sulakshana
Department of Botany, University College for Women, Koti, Hyderabad, Telangana State, India

G. Neelima · R. Mukhopadhyay · K. S. N. Jyothi
Department of Biotechnology, University College for Women, Koti, Hyderabad, Telangana State, India

D. K. Choudhary et al. (eds.), *In Silico Approach for Sustainable Agriculture*,
https://doi.org/10.1007/978-981-13-0347-0_15

15.1 Introduction

Plants are the rich source of metabolites with impressive biological properties. The plant metabolites are organic compounds synthesized using enzyme-mediated chemical reactions called metabolic pathways. Metabolites, synthesized by plants, have both essential and specific functions. The chemical compounds which are essential for growth, development, and reproduction of the plants are known as primary metabolites. These are the important components in maintaining normal physiological processes; thus, it is often referred to as central metabolites. These are produced by common metabolic pathways that are essential to the plant's survival. These metabolites are typically formed during the growth phase as a result of energy metabolism and deemed essential for proper growth. Primary metabolites are found universally in the plant kingdom because they are the constituents or products of fundamental metabolic pathways or cycles such as glycolysis, Krebs cycle, Calvin cycle, etc.

Primary metabolites are essential because the primary pathways enabling a plant to synthesize assimilate and degrade organic compounds. Examples of primary metabolites include energy-rich fuel molecules such as sucrose and starch, structural components such as cellulose, informational molecules such as DNA (deoxyribonucleic acid) and RNA (ribonucleic acid), and pigments, such as chlorophyll. In addition to having fundamental roles in plant growth and development, some primary metabolites are precursors (starting materials) for the synthesis of secondary metabolites. Other examples of primary metabolites include alcohols such as ethanol, lactic acid, and certain amino acids. Within the field of industrial microbiology, alcohol is one of the most common primary metabolites used for large-scale production. Specifically, alcohol is used for processes involving fermentation which produce products like beer and wine. Another example of a primary metabolite commonly used in industrial microbiology includes citric acid. Citric acid, produced by *Aspergillus niger*, is one of the most widely used ingredients in food production. It is commonly used in pharmaceutical and cosmetic industries as well.

15.2 Secondary Metabolites

The plants also produce certain organic compounds through the modification of primary metabolite pathways. These are referred to as secondary metabolites, and they do not play a direct role in growth and development of plant (Fig. 15.1). These are produced during the end or near the stationary phase of growth. Many of the secondary metabolites serve as antibiotics, insecticidal compounds, phytoalexins, and other defense-related chemicals. Some of them function as flavoring, fragrant agents, and coloring pigment. They mediate the interaction of plants with other organisms, such as plant-pollinator, plant-pathogen, and plant-herbivore. These

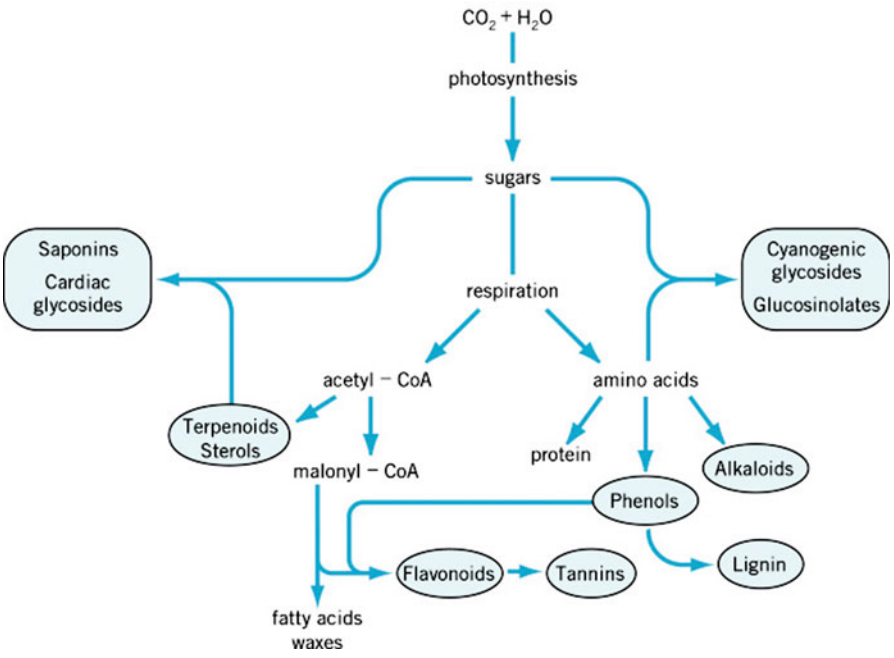


Fig. 15.1 Secondary metabolites are derived from primary metabolites

compounds are variously distributed in the plant kingdom, and their functions are specific to the plants in which they are found.

Secondary metabolites have a great potential application in industry. Some of the examples are atropine and antibiotics such as erythromycin and bacitracin. Atropine, derived from various plants, has important clinical use in treatment of bradycardia. Antibiotics such as erythromycin and bacitracin are other important secondary metabolites. Erythromycin, derived from *Saccharopolyspora erythraea*, is a commonly used antibiotic with a wide antimicrobial spectrum. Bacitracin, derived from *Bacillus subtilis*, is an antibiotic commonly used as a topical drug. Bacitracin is synthesized in nature as a nonribosomal peptide synthetase that can synthesize peptides; however, it is used in the clinic as an antibiotic. Many of the plants develop secondary metabolites, which are pharmacologically active (Samuelsson and Bohlin 2009). These are part of their own protective mechanism against phytopathogens or the result of plant's normal program of growth (*preformed antifungal compounds*) or can be produced in response to pathogenic attack (*induced antifungal compounds or phytoalexins*) (Morrissey and Osbourn 1999).

The secondary metabolites in plants can be divided into different categories according to their biosynthetic principles (Agostini-Costa et al. 2012). All these groups contain compounds with similar biosynthetic properties, and the compounds within the groups do also have some similarities in their structures (Ruby and Rana 2015). A simple classification includes three main groups (Fig. 15.2).

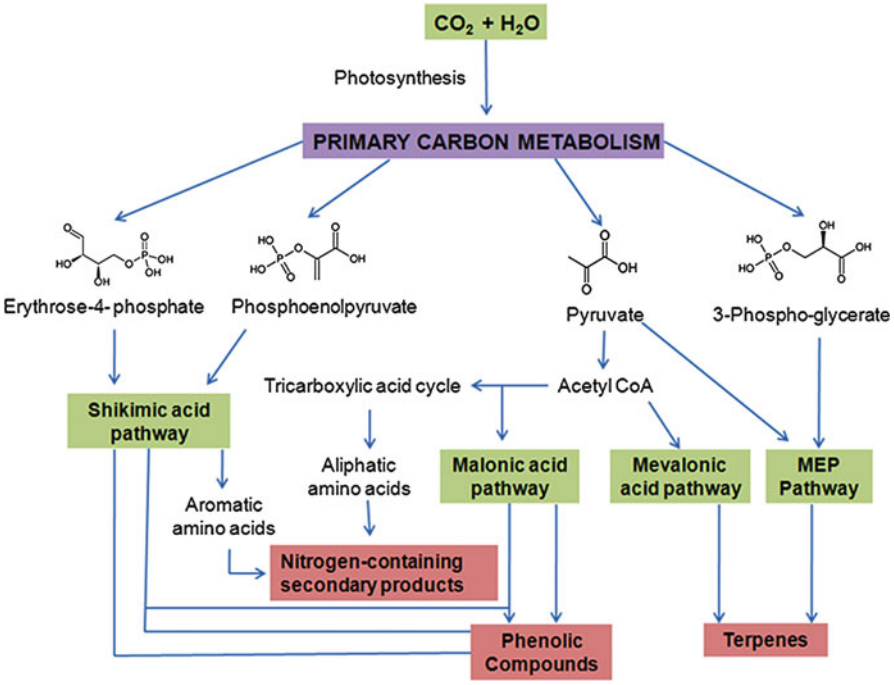


Fig. 15.2 The three major classes of secondary metabolites: nitrogen-containing compounds, phenolic compounds, and terpenes

1. **Nitrogen-containing compounds** such as alkaloids and glucosinolates
2. **Terpenes** such as mono-, di-, tri-, sesqui-, and tetraterpenes, saponins, steroids, cardiac glycosides, and sterols
3. **Phenolic compounds** such as phenolic acids, coumarins, lignans, stilbenes, flavonoids, tannins and lignins

15.3 Plant Secondary Metabolites and Their Biological Properties

Many thousands of secondary metabolites have been isolated from plants, and many of them have powerful physiological effects in humans and are used as medicines. It is only since the late twentieth century that secondary metabolites have been clearly recognized as having important functions in plants. Research has focused on the role of secondary metabolites in plant defense.

15.3.1 Nitrogen-Containing Compounds

They include alkaloids, cyanogenic glucosides, and nonprotein amino acids. Most of them are biosynthesized from common amino acids. All are of considerable interest because of their role in antiherbivore defense and toxicity to predators.

15.3.2 Alkaloids

Alkaloids are a large family of nitrogen-containing secondary metabolites found in approximately 20% of the species of vascular plants, dicots, few monocots, and gymnosperms (Fig. 15.3). Alkaloids are derived from the primary metabolites, amino acids such as tryptophan, tyrosine, aspartic acid, and lysine. Alkaloids accumulate in plant organs such as leaves or fruits; are extremely toxic, especially to mammals; and act as potent nerve poisons, enzyme inhibitors, or membrane transport inhibitors (Wink et al. 1998). The major alkaloid classes include pyrrolidine, tropane, piperidine, pyrrolizidine, quinolizidine, isoquinoline, and indole. Some of the examples for alkaloids are **caffeine**, **morphine** isolated from *Papaver somniferum*, **vincristine** and **vinblastine** from the periwinkle, and **quinine** from the bark of the *Cinchona* tree.

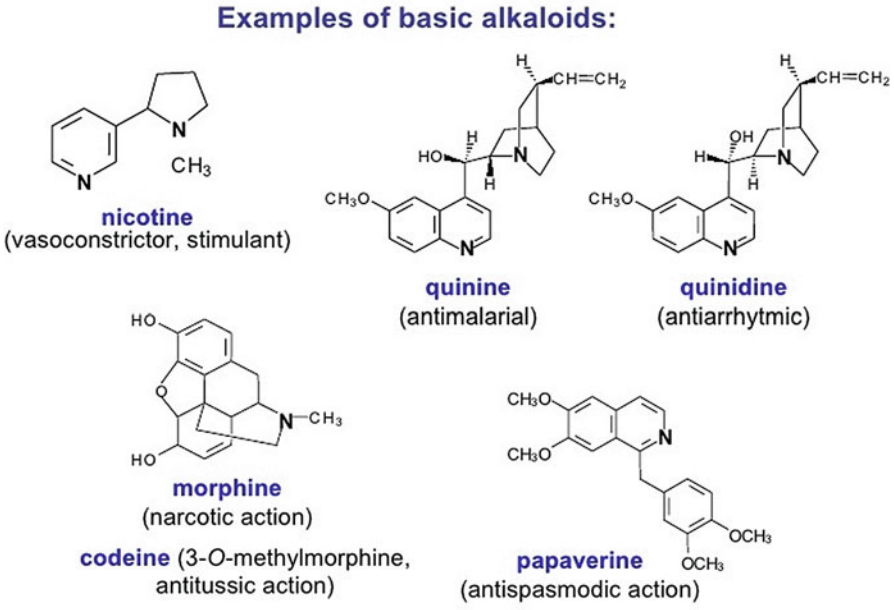


Fig. 15.3 Examples of some basic alkaloids

15.3.3 Cyanogenic Glucosides

Cyanogenic glycosides constitute a group of nitrogen-containing protective compounds other than alkaloids, occurring in members of Gramineae, Rosaceae, and Leguminosae. They are not toxic by themselves, but are readily broken down to give off volatile poisonous substances like HCN and H_2S when the plant is crushed. Their presence deters feeding by insects and other herbivores such as snails and slugs. Examples of these kind of substances include **amygdalin** (from almond, apricot, cherries, and peach), **dhurrin** (from sorghum) (Poulton 1990), and **glucosinolates**, also called as mustard oil glycosides.

15.3.4 Nonprotein Amino Acids

Many plants also contain unusual amino acids called nonprotein amino acids incorporated into proteins but are present as free forms. These act as protective defensive substances. Examples include **canavanine** and **azetidine-2-carboxylic acid** (Fig. 15.4) which are close analogs of arginine and proline, respectively. They elicit their toxicity in various ways; some block the synthesis of proteins or uptake of protein amino acids, while others are incorporated in protein resulting in nonfunctional proteins.

15.3.5 Terpenes or Terpenoids

The terpenes or terpenoids constitute the largest class of secondary products and are united by their common biosynthetic origin from acetyl CoA or glycolytic intermediates. Terpenes are derived from repetitive fusion of five carbon (C_5) isoprene units (Fig. 15.5) joined in a head-to-tail manner and are found in almost all plants (Kogan et al. 2006). Terpenoids are broadly classified on the basis of the number of isoprene units as mono-, di-, tri-, tetra-, sesqui-, polyterpenoids, etc. (Table 15.1): a few typical examples of monoterpenoids, *camphor*, *eucalyptol*, *menthol*, and *thymol*

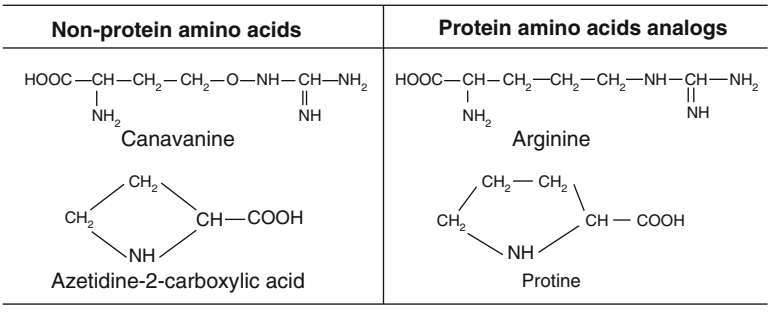


Fig. 15.4 Nonprotein amino acids and their protein amino acid analogs

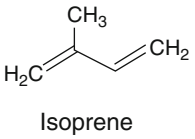


Fig. 15.5 Isoprene unit in terpenes

Table 15.1 Classification of terpenes based on the number of isoprene units with examples

Isoprene units	Carbon atoms		
n	*n*	Name	Example
1	5	Hemiterpenes	Isoprene
2	10	Monoterpenes	Thymol
3	15	Sesquiterpenes	δ-Cadinene
4	20	Diterpenes	Taxol
6	30	Triterpenes	β-Amyrin
8	40	Tetraterpenes	β-Carotene
9–30,000	>40	Polyterpenes	Rubber

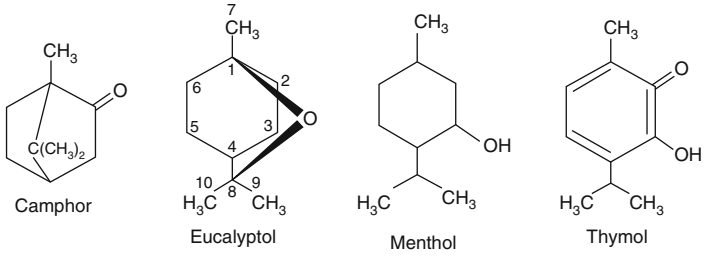


Fig. 15.6 Some examples of monoterpenoids: camphor, eucalyptol, menthol, and thymol

(Fig. 15.6); diterpenoids, paclitaxel (commonly known by the brand name Taxol); triterpenoids, azadirachtin, a complex limnoid; and polyterpenes, rubber (400 isoprene units) isolated from **latex**, from *Hevea brasiliensis.*

15.3.6 Phenolic Compounds

The phenols consist of a **hydroxyl group** (–OH) attached to an aromatic ring. They serve as defense compounds against herbivores and pathogens, attract pollinators, act as fruit dispersers, and absorb harmful UV radiation.

One of the largest classes of plant phenolic compounds is the flavonoids, which perform pigmentation and defense. These are further divided into anthocyanins, flavones, flavonols, and isoflavonoids. *Anthocyanins* range in color from red to blue and purple and are believed to protect against heart disease, diabetes, and cancer. The next two groups flavones and flavonols function to protect cells from UV-B radiation. **Isoflavonoids** have different biological activities such as insecticidal actions, antiestrogenic effects, and anticancer benefits (Park et al. 2001).

Tannins are the second category of plant phenolic polymers which help to bind the collagen protein of animal hides, thereby increasing their resistance to heat, water, and microbes. In general, tannins act as feeding repellents to a great diversity of animals. The other medically relative phenol is salicylic acid, which is the active ingredient in aspirin which has been used to effectively treat aches and fevers. Another type of phenol is **lignin**, which adds stiffness and strength to cell walls of plant cells. Its physical toughness deters feeding by herbivorous animals, and its chemical durability makes it relatively indigestible to herbivores and insects. Lignifications block the growth of pathogens and are a frequent response to infection and wounding.

15.4 Applications of Plant Metabolites

Plant secondary metabolites have many biological and pharmacological properties. They have immense potential as lifesaving drugs, flavor and fragrances, dye and pigments, pesticides, and food additives. Hence, there is a great interest in identification and structure elucidation of medicinal importance plant products in drug development.

Today, large amounts of data related to metabolic reactions are available. But, despite this wealth of information, metabolic phenotypes still cannot be accurately predicted, both in terms of properties of its molecular components (Sweetlove et al. 2008). Knowledge of plant metabolism represents challenges due to the large amount of plant compounds and complexity of their metabolic pathways. The metabolite production in plants is also dependant on strong interaction between the genotype and phenotype and complex regulatory interaction among them. The knowledge on plant metabolite synthesis, accumulation, and storage is important because many of them have the potential for the generation of pharmaceutical products or are key secondary metabolites of commercial interest. They can enhance plant growth or defense against pest attack and can also improve the yield and nutritional quality of crops (nutraceuticals).

The transition of a natural compound from a "screening hit" through a "drug lead" to a "marketed drug" is associated with increasingly challenging demands and coherent efforts from various branches of science. A highly integrated interdisciplinary approach is needed to discover the drugs from the plant origin (Schmidt et al. 2012). The recent technological advances and research trends clearly indicate that natural products are the most important sources of new drugs in the future. Thus, the identification of novel drug targets from the secondary metabolites of plants is as complex as any other natural product research.

15.5 Drug Designing

Chemical space is a concept in cheminformatics and defined as the total descriptor space that encompasses all the small chemical compounds. The chemical space is very large which includes known chemical space (organic molecules reported so far)

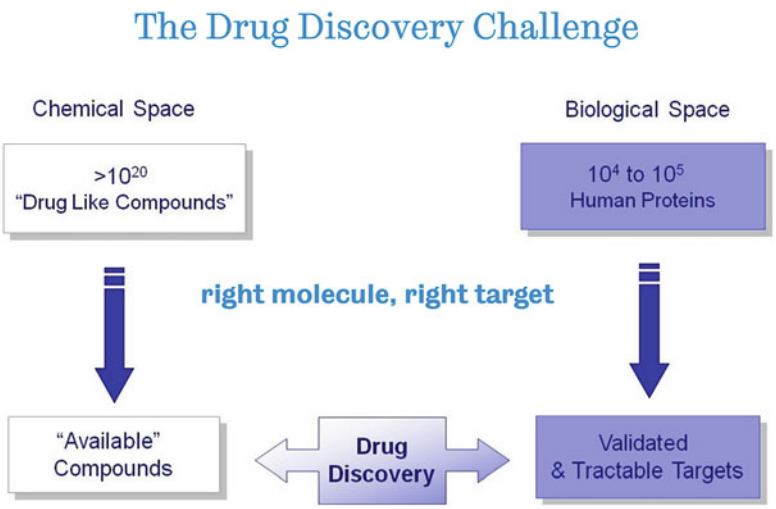


Fig. 15.7 Overview of drug discovery process

and unknown chemical space (novel and uncharacterized molecules). The first step in drug discovery involves screening of this chemical space to choose right molecule (bioactive ligand). The biological space is referred to as a space complementary to that of chemical space (Haggarty 2005). The biological space is also large including numerous proteins that undergo screening to be selected as validated target. Once the right molecules and right target have been chosen, their binding interactions need to be evaluated further in drug discovery pipeline (Fig. 15.7).

15.6 Importance of Natural Products As Lead Structures

Plant secondary metabolites represent an immensely rich source of chemical and structural diversity with an extremely high potential to yield a wealth of lead structures toward many diseases. Only a small percentage, however, roughly 200,000 plant species on earth have been studied chemically. The repository of plant-derived natural products hence deserves to be investigated even more intensely than it has been up to present (Sumner et al. 2003).

Owing to their pharmaceutical and industrial importance, these natural products as well as their biosynthetic mechanisms have been subject of particular interest and extensive characterization. The structural diversity arising from combinatorial complexity of their biosynthesis is the reason why these natural products are a great source of drugs. Understanding the mechanisms of their biosynthesis and devising clever strategies to tweak it can potentially yield fruitful results in the form of economically important products. Linking biosynthetic genes to secondary metabolites and vice versa can potentially help not only in characterization of new

secondary metabolites but also in redesigning known biosynthetic pathways of secondary metabolites to produce novel compounds (Goossens et al. 2003).

15.7 Need for In Silico Characterization of Plant Metabolites

The significant impact of natural products (NPs) on the discovery of therapeutic agents is based on their embedded biosynthetic molecular recognition. Despite the pivotal role of NPs in drug discovery, their use over the past two decades has decreased in the pharmaceutical industry (Newman and Cragg 2012). This unfortunate downturn is mainly attributed to the availability of the materials and the time and cost of isolating and identifying active NPs from extracts (Pascolutti and Quinn 2014). Today, the process of drug discovery has been revolutionized with the advent of genomics, proteomics, bioinformatics, and efficient technologies like combinatorial chemistry, high-throughput screening (HTS), virtual screening, de novo design, in vitro and in silico ADMET screening, and structure-based drug designing. In silico methods can help in identifying drug targets via bioinformatics tools.

They can also be used to analyze the target structures for possible binding/active sites, generate candidate molecules, check for their druglikeness, dock these molecules with the target, rank them according to their binding affinities, and further optimize the molecules to improve binding characteristics (Wadood et al. 2013). The use of computers and computational methods permeates all aspects of drug discovery today and forms the core of structure-based drug design. High-performance computing, data management software, and the Internet are facilitating the access to huge amount of data. This massive complex biological data is transformed into workable knowledge in modern day drug discovery process. The use of complementary experimental and informatics techniques increases the chance of success in many stages of the discovery process. The large amount of data on NPs will help in identification of novel targets, elucidation of their functions, and discovery and development of lead compounds with desired properties.

15.8 In Silico Screening and Characterization of Secondary Metabolites

Computational tools offer the advantage of delivering new drug candidates more quickly and at a lower cost. Major roles of computation in drug discovery are (1) virtual screening and de novo design, (2) in silico ADME/T prediction, and (3) advanced methods for determining protein-ligand binding. Thus, bioinformatics has played an important role in in silico identification of new secondary metabolites.

The genome mining and several pioneering studies have been successful in experimental characterization of new metabolites predicted by in silico analysis. The in silico characterization of plant metabolites involves seven crucial steps.

15.8.1 Screening of Plant Databases for Identification of Uncharacterized Secondary Metabolites

Natural products are among the most important sources of lead molecules for drug discovery. With the development of affordable with whole-genome sequencing technologies and other "omics tools," the field of natural products research is currently undergoing a shift in paradigms. While, for decades, mainly analytical and chemical methods gave access to this group of compounds, nowadays genomics-based methods offer complementary approaches to find, identify, and characterize such molecules. This paradigm shift also resulted in a high demand for computational tools to assist researchers in their daily work (Chen 2011).

A comprehensive and intellectual manual literature mining can be performed for selecting the appropriate secondary metabolite for drug discovery using the following simple analysis of whether or if there is any:

(a) Direct experimental evidence that shows the use of the metabolite
(b) Biological functions of the metabolite
(c) Absence of the metabolite in other plants or organisms
(d) Physicochemical suitability of the compound for designing drugs or inhibitors
(e) Metabolite analogs if any, tested for their therapeutic potential
(f) Efficiency of the molecule to act as a Broad-spectrum target

Presently there are many tools and databases available to mine, identify, and characterize the natural products like secondary metabolites and their biosynthetic enzymes and their genes. There are web portals being developed to provide a one-stop catalog and links to all bioinformatics sources. A prominent example is the Secondary Metabolite Bioinformatics Portal (SMBP) (Weber and Kim 2016). For decades, the only way to get access to new compounds was to cultivate antibiotics-producing microorganisms, mainly fungi and bacteria, under different growth conditions, and then isolate and characterize the compounds with sophisticated analytical chemistry.

Nowadays, "omics" approaches offer complementary access to natural products, by identifying natural product/secondary metabolite biosynthetic gene clusters (BGCs) (Fig. 15.8). It is possible to assess the genetic potential of producer strains and to more effectively identify previously unknown metabolites. While this approach has led to some renaissance of natural product research in academia and industry, this information will also be the basis to rationally engineer molecules or develop "designer molecules" using synthetic biology approaches in the future. Although the diversity of natural product chemical scaffolds is vast, the biosynthetic

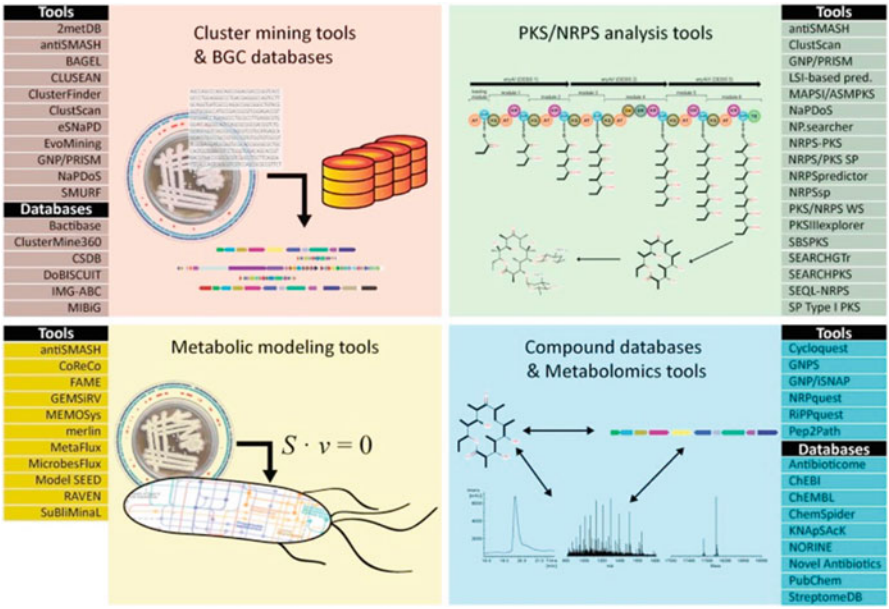


Fig. 15.8 Overview of the most commonly used and freely accessible tools specialized for the analysis of secondary metabolites and their pathways

principles are highly conserved for many secondary metabolites. There is a set of enzyme families, which are often and very specifically associated with the biosynthesis of different classes of secondary metabolites. Thus, sequence information of these known gene families can be used to mine genomes for the presence of secondary metabolite biosynthetic pathways.

There are two principal strategies in the implementation of bioinformatics tools. Rule-based approaches can be used to identify gene clusters encoding known biosynthetic routes with high precision. In the first step of the mining process, these tools identify genes encoding conserved enzymes/protein domains that have associated roles in secondary metabolism, for example, the "condensation (C)," "adenylation (A)," and "peptidyl carrier protein (PCP)" domains of nonribosomal peptide synthetases (NRPSs). In the second step, predefined rules are used to associate the presence of such hits with defined classes of natural products.

15.8.2 Ligand-Based Virtual Screening of Selected Uncharacterized Secondary Metabolites

Virtual screening or in silico screening is a new branch of medicinal chemistry that contributes to a rapid and cost-effective tool for computational screening of

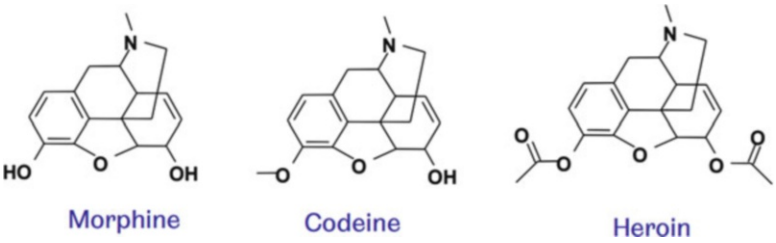


Fig. 15.9 Pain killer drugs

databases in search of novel drug leads. This approach quickens the drug discovery process by short-listing high-quality druglike compounds. Virtual screening is classified into two types. They are ligand-based virtual screening and structure-based screening. Ligand-based virtual screening methods increase the chance of identifying more (and more diverse) active compounds from a virtual screening campaign (Kruger and Evers 2010). In this chapter uncharacterized plant metabolites are considered as ligands which are going to be lead molecules (compounds that have pharmacological and biological activity) in the future. Computational filters should be applied to eliminate undesirable compounds from further considerations.

Virtual screening of small molecule databases via ligand is one of the crucial steps in the initial stages of drug designing process. It is based on the similarity principle, and it minimizes the chemical space of huge databases to a manageable size where selected ligands can be tested experimentally (Supratik Kar, Kunal Roy 2013). The rationale for similarity searching is that structurally similar molecules tend to have similar chemical properties. For example, morphine, codeine, and heroin are the drugs used to relieve pain. Though they have been isolated from different sources, all of them function as pain killers as they are structurally similar and shown in the Fig. 15.9. Thus similar structures have same or similar functions.

The structure similarities between the selected uncharacterized secondary metabolite and characterized bioactive compounds can be identified by using bioinformatics databases and tools. Thus unknown structure (query) is given as input to the tool, and the tool will search against various small molecule databases to give the highest similar structures as best hits. Some of the most popularly used databases and tools for this step are listed in Tables 15.2 and 15.3.

15.8.3 Drug Likeness/Druggability Estimation

Druglikeness is a qualitative concept used in drug design. The phrase "druglike" generally means molecules which contain functional groups and/or have properties consistent with the majority of known drugs. It is predicted from the molecular structure before the substance is even synthesized and tested (Tian et al. 2015). Druglikeness evaluation can be carried out using Lipinski's rule of

Table 15.2 Popular databases for characterization of plant metabolites

S. no.	Databases	Description	URL
1.	PubChem	It is a freely available database of chemical structures of small organic molecules and information on their biological activities. It contains structure, nomenclature, and calculated physicochemical data and is linked with NIH PubMed/Entrez information	https://pubchem.ncbi.nlm.nih.gov/
2.	DrugBank	The database is a blended bioinformatics and cheminformatics resource that combines detailed drug (i.e., chemical, pharmacological, and pharmaceutical) data with comprehensive drug target (i.e., sequence, structure, and pathway) information	https://www.drugbank.ca/
3.	Therapeutic Target Database (TTD)	It is a drug database designed to provide information about the known therapeutic protein and nucleic acid targets	
4.	The Universal Natural Products Database (UNPD)	It provides access to chemical information relevant for virtual activity screening of a large number (>200,000) of natural products	https://omictools.com/universal-natural-products-database
5.	Chemical Entities of Biological Interest (ChEBI)	It is a freely available dictionary of molecular entities focused on "small" chemical compounds	http://www.ebi.ac.uk/chebi/
6.	ChEMBL	It is a publicly available database of drugs, druglike small molecules, and their targets. The database is unique because of its focus on all aspects of drug discovery and its size	https://www.ebi.ac.uk/chembl/
7.	SuperNatural	It is a comprehensive natural product database (>300,000 metabolites) with information about chemical structures, structural relatedness, mechanism of action, metabolite target pairs, and commercial availability	http://bioinf-applied.charite.de/supernatural_new/index.php
8.	The ZINC database	It is a repository generated from records of various databases and can be searched by selecting "natural products" as a subset	http://zinc.docking.org/
9.	In Vivo/In Silico Metabolites Database (IIMDB)	It consists of both known and computationally generated compounds	http://metabolomics.pharm.uconn.edu/iimdb/
10.	TCM database	It contains structural records for more than 20,000 metabolites isolated from traditional Chinese medicines	http://tcm.cmu.edu.tw/

Table 15.3 Important tools used in screening of structure similarity

S. no.	Tools	Description	URL
1.	LiSiCA (Ligand Similarity using Clique Algorithm)	It is a ligand-based virtual screening software that searches for 2D and 3D similarities between a reference compound and a database of target compounds	https://omictools.com/ligand-similarity-using-clique-algorithm-tool
2.	CHAAC	It is a ligand-based virtual screening tool. It compares your molecule with a database of ligands and outputs a list of candidates with similar chemical profile to that of your query	
3.	Swiss Similarity web tool	Web tool for rapid ligand-based virtual screening of small to unprecedented ultralarge libraries of small molecules. Screenable compounds include drugs, bioactive and commercial molecules	http://www.swisssimilarity.ch/
4.	Ultrafast Shape recognition-Virtual Screening (USR-VS)	It is a web server for ligand-based virtual screening powered by ultrafast shape recognition techniques	http://usr.marseille.inserm.fr/
5.	ChemMine web tools	It is an online service for analyzing and clustering small molecules by structural similarities, physicochemical properties, or custom data types	http://chemmine.ucr.edu/
6.	ChemMapper	It is an online platform to predict polypharmacology effect and mode of action for small molecules based on 3D similarity computation	https://omictools.com/chemmapper-tool
7.	ROCS	It is a powerful virtual screening tool which can rapidly identify potentially active compounds by shape comparison	https://www.eyesopen.com/rocs

five, which was formulated by Christopher (Lipinski in 2012) based on the observation that most orally administered drugs are relatively small and lipo philic molecules (Sweetlove et al. 2008).

As per Lipinski's rule, the molecule should satisfy the following criteria:

- Not more than five hydrogen bond donors
- Not more than ten hydrogen bond acceptors
- A molecular mass less than 500 da
- log*P* not greater than 5

Various available tools for druglikeness are listed in Table 15.4.

Table 15.4 Tools for evaluating druglikeness

S. no.	Tool	Description	URL
1.	DruLiTo	It is an open-source virtual screening tool. Its calculation is based on the various druglikeness rules like Lipinski's rule, MDDR-like rule, Veber rule, Ghose filter, BBB rule, CMC-50 like rule, and quantitative estimate of druglikeness (QED)	http://www.niper.gov.in/pi_dev_tools/DruLiToWeb/DruLiTo_index.html
2.	DrugLogit	Predicting the probability of a compound being classified as a drug or nondrug, as well as disease category (or organ) classification (DC)	http://www.vls3d.com/links/chemoinformatics/admet/admet-and-physchem-predictions-and-related-tools
3.	ALOGPS-	It predicts logP values online	http://www.vls3d.com/links/chemoinformatics/admet/admet-and-physchem-predictions-and-related-tools
4.	DrugMint	It is a web server developed for predicting drug likelihood of a compound	http://crdd.osdd.net/oscadd/drugmint/
5.	Molinspiration	It calculates molecular properties and bioactivity score	http://www.molinspiration.com/cgi-bin/properties

15.8.4 ADMET Estimation

ADMET stands for absorption, distribution, metabolism, excretion, and toxicity. ADME is an essential concept that describes potential impact a compound may have on a living system. This is because movement and metabolism of molecules is determined by physicochemical properties of the molecule as well as the host system "kinetics" or "pharmacokinetics". It is defined as movement of molecules and chemical properties such as polarity, molecular size, molecular weight, chirality, and many more that have an effect on the ADME potential of a compound. The prediction of the ADMET properties plays an important role in the drug design process because these properties account for the failure of about 60% of all drugs in the clinical phases. Recent studies have shown that poor pharmacokinetics and toxicity are the major causes of high attrition rates in drug development. Hence, these areas should be considered as early as possible in the drug discovery process, thus improving the efficiency and cost-effectiveness of the industry. Resolving the pharmacokinetic and toxicological properties of drug candidates remains a key challenge for drug developers. Some of the ADME properties evaluated using in silico models are intestinal permeability, human intestinal absorption, human oral bioavailability, aqueous solubility, active transport, blood-brain barrier permeation, efflux by P-glycoprotein, plasma protein binding, metabolic stability, interactions with cytochrome P450s, and toxicity. Table 15.5 enlists the tools for ADMET estimation.

Table 15.5 Tools for ADMET estimation

S. no.	Tool	Description	URL
1.	QikProp	Software that provides rapid ADME predictions of drug candidates. Distributed by Schrodinger	https://www.schrodinger.com/
2.	ADMET predictor	Software for advanced predictive modeling of ADMET properties. ADMET predictor estimates a number of ADMET properties from molecular structures	https://www.simulations-plus.com/software/admetpredictor/
3.	Discovery studio ADMET software	The ADMET collection provides components that calculate predicted absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties for collections of molecules. Distributed by Accelrys	http://accelrys.com/products/collaborative-science/biovia-discovery-studio/qsar-admet-and-predictive-toxicology.html
4.	C2-ADME	It provides fast computation by only requiring 2D structures as input and processes as many as 2 million molecules/h	www.accelrys.com
5.	TOPKAT	Cross-validated models for the assessments of chemical toxicity from chemical's molecular structure	www.accelrys.com
6.	GastroPlus	It simulates the oral absorption, pharmacokinetics, and pharmacodynamics for drugs in human and preclinical species	www.simulations-plus.com

15.8.5 Target Prediction

Molecular insight into the mode of action of bioactive small molecules is key to understanding observed phenotypes, predicting potential side effects or cross-reactivity, and optimizing existing compounds. In particular, mapping their targets is a crucial step toward providing a rational understanding of small molecule's bioactivity. For these reasons, high-throughput reverse screening of chemical compounds against arrays of protein targets has become an integral part of drug discovery pipelines (Keiser et al. 2007). A knowledge-based approach to computationally identify new targets for uncharacterized molecules has been facilitated by large datasets of protein-small molecule interactions. Computational predictions play an important role in narrowing down the set of potential targets. In particular, the large amount of information collected on protein-small molecule interactions in the last few years has enabled researchers to develop ligand-based approaches for target prediction (Ziegler et al. 2013). The Table 15.6 gives a description of the tools for target prediction.

Table 15.6 Tools for target prediction

S. no.	Tool	Description	URL
1.	SwissTargetPrediction	Online tool to predict the targets of bioactive small molecules in human and other vertebrates. This is useful to understand the molecular mechanisms underlying a given phenotype or bioactivity, to rationalize possible side effects	http://www.swisstargetprediction.ch/
2.	CSNAP (Chemical Similarity Network Analysis Pull-down)	A network-based approach for automated compound-target identification	http://services.mbi.ucla.edu/CSNAP/
3.	SuperPred	It is a prediction webserver for target prediction of compounds	http://prediction.charite.de/
4.	TargetHunter of Small Molecule	It is a web portal to identify possible targets of small molecules by searching the available bioactive compound-target pairs reported from literature using the query structure	http://www.cbligand.org/TargetHunter/
5.	HitPick web server	It facilitates the analysis of chemical screenings by identifying hits and predicting their molecular targets	http://mips.helmholtz-muenchen.de/hitpick/cgibin/index.cgi?content=targetPrediction.html

15.8.6 Docking

Molecular docking is a method of modeling in bioinformatics involving the interaction of two or more molecules to give a stable complex. Knowledge of the relative orientations of two interacting partners enables the prediction of the three-dimensional structure of any complex. Molecular docking results in different possible adduct structures and predicts optimized docked conformer that has minimum free energy of overall system (Shoichet et al. 2002). Molecular docking is carried out prior to experimental investigation to predict the feasibility of biochemical reaction particularly interaction between small molecules (ligand) and protein target (may be an enzyme). This information may provide a raw material for the rational drug designing. As molecular docking can predict different binding modes of ligand in the groove of target molecule, more potent, selective, and efficient drug candidates can be developed with ease (Dar and Mir 2017). The tools for docking are listed in Table 15.7.

Table 15.7 Tools for docking

S. no.	Tool	Description	URL
1.	Autodock	It is a suite of automated docking tools. It is designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure	http://autodock.scripps.edu/
2.	DOCK Anchor-and-Grow-based docking program	Free for academic usage. Flexible ligand. Flexible protein. Maintained by the Shoichet group at the UCSF	http://dock.compbio.ucsf.edu/
3.	GOLD GA-based docking program	Flexible ligand. Partial flexibility for protein	https://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/gold/
4.	SwissDock	It is a web service to predict the molecular interactions that may occur between a target protein and a small molecule	http://www.swissdock.ch/
5.	FlexX	It is a software package to predict protein-ligand interactions	https://www.zbh.uni-hamburg.de/en/research/research-group-for-computational-molecular-design/software-server/flexx-molecular-docking.html
6.	Surflex	Fully automatic flexible molecular docking using a molecular similarity-based search engine	https://omictools.com/surflex-dock-tool

15.8.7 Binding Free Energy Estimation

Docking programs create numerous potential ligand poses; some of them can be immediately rejected due to clashes with the protein and the remaining is evaluated using some scoring function. Docking takes a pose as input and gives the likelihood in numerical suggesting that the pose represents a favorable binding interaction and ranks one ligand relative to another.

Most scoring functions are based on molecular mechanics force fields that approximately calculate the energy of the pose within the binding site. The binding free energy can be estimated by the following equation:

$$\Delta G_{\text{bind}} = \Delta G_{\text{solvent}} + \Delta G_{\text{conf}} + \Delta G_{\text{int}} + \Delta G_{\text{rot}} + G_{\text{t/t}} + \Delta G_{\text{vib}}$$

$\Delta G_{\text{solvent}}$ is free energy due to solvent effects.
ΔG_{conf} is free energy due to conformational changes in the protein and ligand.
ΔG_{int} is free energy due to protein-ligand interactions.
ΔG_{rot} is free energy due to internal rotations.
$\Delta G_{\text{t/t}}$ is free energy due to association energy of ligand and receptor to form a single complex.
ΔG_{vib} is free energy due to changes in vibrational modes.

Table 15.8 Tools for binding free energy estimation

S. no.	Tool	Description	URL
1.	Hyde	Entirely new approach to assess binding affinities and contributions to binding of a complex, with a visual feedback at atomic detail	https://www.biosolveit.de/Hyde/
2.	X-score	Program for computing the binding affinities of the given ligand molecules to their target protein	http://sw16.im.med.umich.edu/software/xtool/
3.	NNScore Python script	Used for computing binding free energies from PDBQT files of the receptor and the ligand, using a neural network approach	http://rocce-vm0.ucsd.edu/data/sw/hosted/nnscore/
4.	Binding Affinity Prediction of Protein-Ligand (BAPPL) server	It computes the binding free energy of a non-metallo protein-ligand complex using an all-atom energy-based empirical scoring function	http://www.scfbio-iitd.res.in/software/drugdesign/bappl.jsp

A low (negative) energy designates a stable system and thus a likely binding interaction. Accurate prediction of ligand binding mode and affinity to a target receptor is essential for rational drug design with optimized sensitivity and specificity (Irwin and Shoichet 2016). Therefore, various computational tools have become crucial components in drug discovery projects. The tools for binding free energy estimation are given in Table 15.8.

The various steps involved *in silico* identification and characterization of plant metabolites are represented in the form of flow chart in Fig. 15.10.

15.9 Conclusion

The escalating faith in plant-based medication systems and the extensive research on secondary metabolites of plants are giving a new hope for the development of natural drugs for alleviating human diseases. The faster approach toward scientific data validation and in silico characterization of secondary metabolites has enhanced the quality of research outcome in the arena of drug discovery. The advent of high-throughput technology and ever-growing number of computational tools and databases provides new possibilities for identifying and characterizing novel secondary metabolites. In silico screening and characterization of plant metabolites have the potential to speed the rate of drug discovery while reducing the need for expensive lab work and clinical trials. The in silico characterized plant metabolites might be good candidates as drugs for various therapies. Further in vitro validation is required for these compounds to provide insights into their mode of action against the

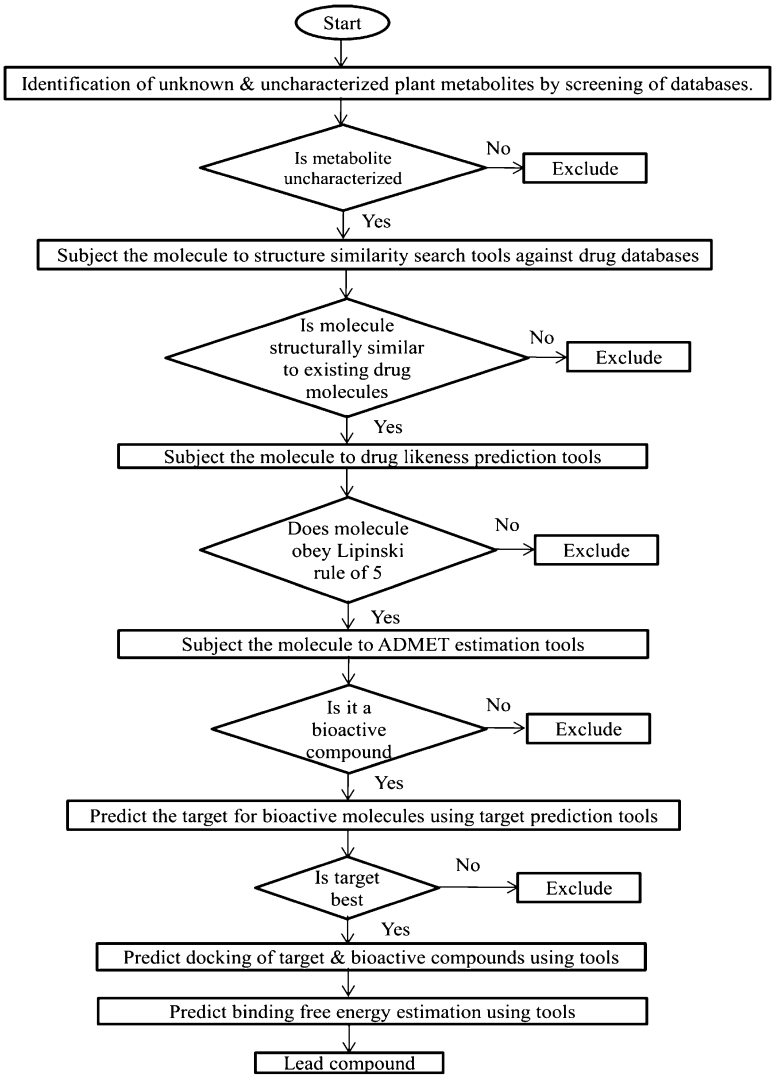


Fig. 15.10 Schematic illustration of in silico identification and characterization of plant metabolites as lead molecules via ligand-based virtual screening approach

different diseases. In summary, in silico analysis suggests that the predicted compounds from plants offer an attractive starting point and a broader scope to mine for potential drugs.

References

Agostini-Costa TS, Vieira RF, Bizzo HR, Silveira D, Gimenes MA (2012) Secondary metabolites. In: Dhanarasu DS (ed) Chromatography and its applications. In Tech, Brazil

Chen CY (2011) TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico, 2011. PLoS One 6:e15939. https://doi.org/10.1371/journal.pone.0015939

Cheng F, Li W, Liu G, Tang Y (2013) In silico ADMET prediction: recent advances, current challenges and future trends. Curr Top Med Chem 13(11):1273–1289., 1568-0266/1873-4294. https://doi.org/10.2174/15680266113139990033

Dar AM, Mir S (2017) Molecular docking: approaches, types, applications and basic challenges. J Anal Bioanal Tech. 2017 8:2. https://doi.org/10.4172/2155-9872.1000356

Feixiong C, Weihua L, Guixia L, Yun T (2013) In Silico ADMET prediction: recent advances, current challenges and future trends. Curr Top Med Chem 13:1273–1289

Goossens A, Hakkinen ST, Laakso I, Seppanen-Laakso T, Biondi S, De Sutter V (2003) A functional genomics approach toward the understanding of secondary metabolism in plant cells. Proc Natl Acad Sci U S A 100:8595–8600

Haggarty SJ (2005) The principle of complementarity: chemical versus biological space. Curr Opin Chem Biol 9(3):296–303

Irwin JJ, Shoichet BK (2016) Docking screens for novel ligands conferring new biology. J Med Chem 59:4103–4120

Kabera JN, Seman1 E, Mussa AR, He X (2014) Plant secondary metabolites: biosynthesis, classification, function and pharmacological classification, function and pharmacological properties. J Pharm Pharmacol 2(7):377–392

Kar S, Roy K (2013) How far can virtual screening take us in drug discovery? Expert Opin Drug Discovery 8(3):245–261. https://doi.org/10.1517/17460441.2013.761204

Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK (2007) Relating protein pharmacology by ligand chemistry. Nat Biotechnol 25:197–206

Kogan SB, Kaliya M, Froumin N (2006) Liquid phase isomerization of isoprenol into prenol in hydrogen environment. Appl Catal A Gen 297(2):231–236

Kristensen TG, Nielsen J, Pedersen CNS (2013) Methods for similarity-based virtual screening. Comput Struct Biotechnol J 5:e201302009

Kruger DM, Evers A (2010) Comparison of structure- and ligand-based virtual screening protocols considering hit list complementarity and enrichment factors. ChemMedChem 5:148–158

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2012) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 64:4–17

Morrissey JP, Osbourn AE (1999) Fungal resistance to plant antibiotics as a mechanism of pathogenesis. Microbiol Mol Biol Rev 63:708–724

Newman DJ, Cragg GM (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. J Nat Prod 75:311–335

Osbourn AE, lanzotti V (2009) Plant-derived natural products- synthesis, function and application. Springer, LLC, New York

Park ES, Moon WS, Song MJ, Kim MN, Chung KH, Yoon JS (2001) Antimicrobial activity of phenol and benzoic acid derivatives. Int Biodeterior Biodegrad 47(4):209–214

Pascolutti M, Quinn RJ (2014) Natural products as lead structures: chemical transformations to create lead-like libraries

Poulton JE (1990) Cyanogenesis in plants. Plant Physiol 94(2):401–405

Reymond J-L, Awale M (2012) Exploring chemical space for drug discovery using the chemical universe database. ACS Chem Neurosci *3*(9):649–657

Ruby T, Rana CS (2015) Plant secondary metabolites: a review. Int J Eng Res Gen Sci 3 (5):661–670

Samuelsson G, Bohlin L (2009) Drugs of natural origin, 6th edn. Apotekarsocieteten, Sweden

Schmidt TJ, Khalid SA, Romanha AJ, Alves TM, Biavatti MW, Brun R, Da Costa FB, de Castro SL, Ferreira VF, de Lacerda MV (2012) The potential of secondary metabolites from plants as drugs or leads against protozoan neglected diseases – part I. Curr Med Chem 19(14):2128–2175

Shoichet BK, McGovern SL, Wei B, Irwin JJ (2002) Lead discovery using molecular docking. Curr Opin Chem Biol 6:439–446

Sumner LW, Mendes P, Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry 62:817–836

Sweetlove LJ, Fell DA, Fernie AR (2008) Getting to grips with the plant metabolic network. Biochem J 409:27–41

Tian S, Wang J, Li Y, Li D, Xu L, Hou T (2015) The application of in silico drug-likeness predictions in pharmaceutical research. Adv Drug Deliv Rev 86:2–10

Wadood A, Ahmed N, Shah L, Ahmad A, Hassan H, Shams S (2013) In-silico drug design: an approach which revolutionarised the drug discovery process. OA Drug Des Deliv. 2013 Sep 01 1(1):3

Weber T, Kim HU (2016) The secondary metabolite bioinformatics portal: computational tools to facilitate synthetic biology of secondary metabolite production. Synth Syst Biotechnol 1:69–79. https://doi.org/10.1016/j.synbio.2015.1012.100

Wink MT, Schmeller TB, Latz Bruning B (1998) Modes of action of allelochemical alkaloids: interaction with neuroreceptors, DNA and other molecular targets. J Chem Ecol 24 (11):1881–1937

Ziegler S, Pries V, Hedberg C, Waldmann H (2013) Target identification for small bioactive molecules: finding the needle in the haystack. Angew Chem Int Ed Eng 52:2744

Chapter 16
In Silico Identification of Plant-Derived Secondary Metabolites in Defense

Kalaivani K. Nadarajah

Abstract Plant secondary metabolites contribute significantly to the field of agriculture, medicine, and biofuels. These compounds have been a focal point in plant breeding and metabolic engineering. However information on these compounds is lacking severely especially in non-model plants. Through integrated omics approach, we can now study secondary metabolites in model and non-model plants to determine genes, predict gene function, and provide information on pathways that may regulate its biosynthesis and function. Online resources have provided a means to fast-track our understanding on the mechanism involved in the biosynthesis of secondary metabolites and how these products are regulated by their environment, developmental stages, and species. The information derived may be utilized in metabolic engineering or in elicitation of the mechanisms involved in its production. As secondary metabolites have been implicated in plant defense mechanism, the understanding of the genes, their function, and their pathways will definitely assist in improving plant defenses against biotic and abiotic stresses. Here we provide a brief overview on the databases and resources available to conduct in silico analysis of plant secondary metabolites and future prospects in utilizing the derived information to improve metabolite function and production in crops.

Keywords Secondary metabolites · In silico · Plants · Genomics · Proteomics · Transcriptomics · Defense and metabolomics

16.1 Introduction

Expanding populace, deficiency of arable land, the subsequent developing interest for sustenance, and crude materials are significant factors that dictate the direction in plant improvement especially with regards to better protection from biotic and

K. K. Nadarajah (✉)
School of Environmental and Natural Resource Sciences, Faculty and Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia
e-mail: vani@ukm.edu.my

D. K. Choudhary et al. (eds.), *In Silico Approach for Sustainable Agriculture*,
https://doi.org/10.1007/978-981-13-0347-0_16

abiotic stresses (Yan and Kerr 2002) and improvement of yield. What's more, plants assume a focal point for the creation of bio-economy (Dyer et al. 2008) through generating therapeutics, inexhaustible chemicals, and biofuel, simply through basic plant physiology and biology (Yan and Kerr 2002; Saha et al. 2011). Plant science has endeavored to provide answers to some overwhelming issues that face the twenty-first century. Among them are the reduction of accessible nourishment, the exhaustion of worldwide oil reserves, and a mounting shortage of freshwater. Food shortage remains our number one concern. Food security is threatened by losses, hence making defenses against biotic and abiotic elements crucial in addressing food security. As such, studying the elements that contribute toward plant defense and resilience is of economic importance. Here we will look into the platforms and resources that are available to study the metabolites that are responsible to fulfill the defense protocol within plants (Beckers et al. 2016).

Metabolites in plants are divided into primary and secondary metabolites (Fiehn et al. 2000). Primary metabolites are implicated in the developmental and metabolism processes such as the production of carbohydrates, proteins, and lipids. The secondary products however are recognized as the end products of the primary metabolic pathways. These compounds essentially are involved in the production of phenolics, steroids, alkaloid, sterols, etc. which are largely part of the plant's defense arsenal. Secondary metabolites or secondary compounds are chemicals that are not required for normal growth and development and are not made through metabolic pathways common to all plants. In examining the plant kingdom, we observe that the production of secondary metabolites may be different and dependent on the taxonomic group of the plant. One major observation on secondary metabolites in plants is that though they may play an important role in plant defense, they constitute a very small amount of plant metabolites and make up a minute percentage of the plant's dry weight. The small quantities and the synthesis within certain specialized cells further compound the difficulty in extracting and analyzing these chemicals. In this chapter we will focus on the secondary metabolites required for plant defense and how these substances may be analyzed via in silico techniques (Beckers et al. 2016; Alagna 2013).

Our limited knowledge and comprehension of the genetics, mechanisms of these chemicals, and the insufficient repository of information available to us on these secondary metabolites has hampered the advances in metabolic engineering and crop improvement (Junker 2014; Shachar-Hill 2013; Cusido et al. 2014; Sweetlove et al. 2003). Herein lies our need to acquire strategies and technologies that will enable us to garner a large amount of information in the shortest time possible with regards to plant systems and mechanisms (Shachar-Hill 2013; Cusido et al. 2014; Lotz et al. 2014). One promising way to deal with this is through the use of in silico metabolic modeling. Right off the bat, in silico metabolic modeling has demonstrated stunning results in studying metabolic engineering of other systems such as microbes and organisms (Poblete-Castro et al. 2013; Kim et al. 2014; Hwang et al. 2014). Besides, more than 50 plant genomes, mostly crops, have been sequenced (Michael and Jackson 2013) generating bountiful data that is useful in secondary metabolism

modeling (de Oliveira Dal'Molin et al. 2010; Saha et al. 2011). Furthermore, various powerful tools and approaches have been developed to assist us with the study of the genes, gene function, and metabolites in any given plant system (Dersch et al. 2016).

The availability of large computerized data on various plant genomes has encouraged the utilization of bioinformatics tools in the analysis and determination of gene, gene function, gene expression, and cell metabolism (Dersch et al. 2016). The availability of these genome resources is a quick and useful way of analyzing metabolites involved in plant defense against biotic and abiotic stresses. Major classes of plant defense metabolites include thionins, defensins, cyclotides, lipid-exchange proteins, hevein-like, and snakins (as per their homology in plant biodiversity) (Nawrot et al. 2014). Despite the fact that many proteins with a function in plant defense have been identified, a great extent of this information seems unexplored for various applications in agriculture and human needs. Not surprisingly, therefore, that there has been a targeted approach in studying plant defense components and antimicrobial prospects through in silico technologies and platforms (Pestana-Calsa et al. 2010). These genes that are identified and linked to plant defense should be further elucidated to determine the role they play in defense. Following the initial in silico analyses of these targets, candidates may be ascertained for use in crop resistance breeding and transgenics (Terras et al. 1995).

Therefore we should endeavor towards generating designer crops. To do this, it is important to obtain complete information of how plants work within a brief timeframe. We should build our understanding of how plants should function in order to reliably predict how the plant systems will respond in any given situation. Plant metabolites play a crucial role in many biological processes within the plant. Fortunately, as we venture into the twenty-first century, technologies are advancing, and the possibility of acquiring information rapidly is becoming more attainable. A decade ago brought tremendous advances in genomic, transcriptomic, and proteomic datasets that can be used in various applications including metabolite prospection. Over the past decade, more and more information and literature are available that cover issues concerning plant secondary metabolites. However for the bioinformatics tools to be of optimum use, it would require sufficient datasets that have been developed for model species that are accessible for heterologous studies of non-model species. Additionally, bioinformatics tools are trained for use to model species where omics datasets are available. Therefore the existing resources should also support heterology-based studies for non-model species. Therefore it is crucial to have integrated, high-quality data that is useful for application in bioinformatic approaches in screening genomic, transcriptomic, proteomic, and metabolomics data to determine gene sets, functional analysis, systems biology, and metabolic pathways of any given cultivated and wild plant species (Pestana-Calsa et al. 2010).

In this chapter we present the importance of studying plant secondary metabolites and how in silico technologies and platforms may be utilized to conduct omics analysis to identify components involved in plant secondary metabolite pathways. However large-scale datasets must be made available for the omics tools to be utilized efficiently to generate knowledge and information that may be applied to

the agricultural industry. The tools and platforms presented here are suitable for use on all plant secondary metabolites, including those specific to defense response.

16.1.1 Importance of Plant Secondary Metabolites

Enhancing the qualities identified with plant secondary metabolism such as biotic and abiotic defenses as well as plant fitness is the principle focus of the agricultural industry (Pereira et al. 2012; Zhao 2007). In addition to defense response, plant secondary metabolites are also responsible for plant physiological processes like growth and development (Broun 2004, 2005). Although we recognize the importance of this group of metabolites, there is a vast amount of information that is lagging with regard to the genes, their regulation, accumulation, and degradation (Forster et al. 2003). Thus far the studies on plant secondary metabolites have been directed towards mapping individual pathways and intermediates involved in the production of these products. Plant secondary metabolites are regulated from the genome all the way to the metabolome. They work together concertedly as a system. Hence changes at the genome level will effect change to the transcripts which eventually results in change to the metabolites. These changes result in an overall change to various processes within the plant. However changes at each level though correlated are not linear. Thus a proper analysis through an integrated approach of the omics technologies will definitely enhance our understanding of various phytochemicals in model and non-model species (Ulrich-Merzenich et al. 2007; Metlen et al. 2009; Forester and Waterhouse 2009). It is important nevertheless to acquire useful tools which will enable us to analyze these information at a rapid speed and to also apply the currently available information in understanding non-model plant systems (Hammami et al. 2009). The following segments will address how the tools and resources may be used to analyze secondary metabolites. The online resources mentioned are for general analysis of plant substances with no specifics to plant secondary metabolites for defense. The resources are all encompassing.

16.1.2 In Silico Application in Analyzing Secondary Metabolites

The next-generation sequencing technologies (NGS – next-age sequencing) that have been developed over the recent years enables delivery of cost-effective, quality sequence datasets (Metzker 2010). Meanwhile the advances in bioinformatics tools at software and hardware levels have enabled improved managements and data mining of large datasets. Besides NGS providing sequence datasets at the genome level, a more recent technology, RNA-Seq, enables efficient transcriptome analysis by means of profound sequencing advancements. This system provides a

significantly more exact estimation of the level of quality transcripts and their isoforms than any other strategies (Wang et al. 2006). Based on the availability of both genome and transcriptome databases for research, the scientific community is now referring to the twenty-first century as the "post-genomic era." This of course does not mean that we have to stop sequencing genomes, but rather that we should look into the biological significance of the data that has been derived and to also advance the technologies and tools to circumvent all the obstacles that are currently faced to make quick analysis of the available datasets efficient, accurate, and reliable.

More than 30 plant genomes have been completely sequenced (http://www.phytozome.net/) to date. What's more, a few activities focus on genome sequencing of an expansive number of firmly related plants and strains of related species. For example, the SOL-100 task was to sequence the succession of 100 genomes from the Solanaceae family (http://solgenomics.net/lifeform/sol100/see). Conversely, the 1001 Genomes Venture was a means to find the genome variation in the 1001 *Arabidopsis* accessions (http://www.1001genomes.org/). With the fast-accumulating plant genome information, unparalleled opportunities have been presented to inspect the genomic premise of secondary metabolism. Such examination may prompt the identification of genealogically specific genes or metabolites. For instance, a current examination of the genome grouping of a non-seed plant *Selaginella moellendorffii* uncovered that its genome contained microbial terpene synthases, which has never been classified in plants (Li et al. 2012; Schatz et al. 2012). Comparative investigation may help uncover the evolution and diversity of these products based on strain specificity. What's more, the investigation of the genomic premise of secondary metabolites may control directly the course taken by experimental studies in the laboratory.

Though we have close to half a million plant species on the planet, only a very small percentage of these have been sequenced (Hostettman and Terreaux 2000). Each species that is sequenced on the other hand would generate thousands of genes. In *Arabidopsis thaliana* alone, it has been purported to contain around 28,000 genes (Schoof et al. 2004). However most of these plant species have not been sequenced, and in most cases, the genes have not been completely identified. Daily we have thousands of new DNA sequences and databases that are submitted to the GenBank (http://www.ncbi.nlm.nih.gov/genbank/measurements). Because of the quantity of plants and the number of genes in question, it is therefore necessary for high-throughput platforms to be utilized to screen and isolate the genes that are of importance to researchers, in this case the genes regulating secondary metabolite production. In any case, secondary metabolism is a multifarious process that is controlled by genetic-environmental interactions. As a complex process, the production of these metabolites is regulated by many genes. Identifying these genes, identifying their function, and placing them within respective pathways are crucial steps in unveiling the mystery behind the synthesis of these metabolites (Yang et al. 2014).

Genome and transcript dataset searches have proven to be an effective and efficient means of identifying candidate genes that are responsible for plant secondary metabolite production due to the excellent coverage of the organisms' genetic potential in these searches. The in silico technology is a quick means of

prospecting and arriving at the potential genes responsible for secondary metabolites in any given plant species, although further biological validation may be required for confirmation. As an example in both model organisms and in most green plants, the genome has been reported to harbor anything between 20 and 60 defensin genes. However, it has been stipulated that these figures may be an underrepresentation of this group of genes in plant species. Hence it has been recommended that a combinatorial study comparing and contrasting the plant genome database against other resources such genome-wide expressed sequence tag (EST) libraries (Silverstein et al. 2005, 2007) of other plant species would definitely increase the number of detected genes significantly. For example, in *Arabidopsis thaliana* over 300 defensin-like genes were identified through the assignment by hidden Markov model of an in silico search (Manners 2007; Silverstein et al. 2005).

16.1.3 Discovery of Secondary Metabolite Genes

As mentioned above the advent of databases and the development of new online analytical resources have accelerated the process of gene identification, especially the identification of novel genes. In a recent report by Field and Osbourn (2008), the effectiveness of these online resources in prediction of gene clusters in plants has been shown. The large number of databases and genome information flooding the GenBank (http://www.ncbi.nlm.nih.gov/genbank/statistics) has posed a challenge in the identification of gene and gene function. Identification of gene and gene function based solely on sequence homology can only provide a hint on its putative functions. One good example of the challenge faced can be seen in the model plant *Arabidopsis thaliana*. Through the use of sequence homology alone, about 30,000 genes have had their functions annotated based on the similarities shown by their sequences. Further only slightly over 10% of these genes have been functionally confirmed through experimentation (Field et al. 2011; Saito et al. 2008).

A powerful way to deal with gene prediction of secondary metabolites within a metabolic pathway is through co-expression analysis. Genes that are within a particular metabolic pathway can have genes that are co-regulated and thus co-expressed under a single or shared control of this pathway. Therefore if there is a gene within this pathway that has no known function, hence based on the presence and the shared regulation of this gene with others in the pathway may be used to predict a potential role for the gene based on the known function and genes within the pathway (Saito et al. 2008). Information of gene expression data may be utilized to identify gene clusters of co-expressed genes within the predicted function of secondary metabolism. RNA-seq and microarray can be utilized to analyze co-expression data. In addition, this data may be analyzed against other already existing publicly available databases (Naoumkina et al. 2010). Additionally when functional characterization was carried out on the identified genes, the outcome confirmed the efficiency of this approach (Carelli et al. 2011). Through joint analysis of genome assembly and expression data, a large number of candidate gene clusters

were predicted (Grafahrend-Belau et al. 2009). Combining techniques as above is a great way of optimizing the information obtained on genes, gene clusters, and also pathways (Saito et al. 2008; Urbanczyk-Wochniak et al. 2005).

16.1.4 Importance of Predicting Pathways

Tracer techniques have been extensively utilized to outline the biosynthetic pathways of almost all crucial plant secondary metabolites (Hartmann 2007). However, while pathways have been outlined, most of these pathways have segments or steps that remain unknown especially downstream of the secondary product biosynthesis. Knowing details of the pathway will provide the necessary information for engineering of the pathway if and when needed to increase the yield or production downstream (Verpoorte and Memelink 2002).

One excellent example of the abovementioned is the identification of the pathways involved in terpenoid production. Terpenoids are a large and diverse class of natural compounds found in plants, and to date over 40,000 terpenoids have been identified. From all the studies conducted on this class of compounds, two biosynthetic pathways have been identified, and the downstream portion of these pathways is still elusive (Roberts 2007; Lichtenthaler 2000; Rohmer 1999) where only minimal information is available on the biosynthetic steps and their intermediate details. Through the utilization of functional genomic tools, the upstream, downstream, and intermediate components of pathways may be identified. This technique does not only identify the genes involved in the biosynthetic pathway but is also able to determine the origin and the destination of the metabolite produced (Tohge et al. 2005).

16.1.5 The Integration of Genomics, Proteomics, and Metabolomics in Understanding Secondary Metabolites

In order to understand a system like secondary metabolism, it would require the study of the interaction between the genes, proteins, transcripts, and metabolites. This network has been studied through the various omics platforms where there is a combination of genome-transcriptome, transcriptome-proteome, and transcriptome-metabolome in deciphering the complexity of the network. The integrated use of these platforms has been reviewed previously (Fiehn 2001; Fiehn and Weckwerth 2003). BL-SOM is a widely used approach in the analysis of the integrated systems. This approach allows for genes and metabolites that are regulated by the same systems to be clustered, whereby the gene-to-gene and gene-to-metabolite network may be elucidated (Hirai et al. 2005; Mercke et al. 2004). Rischer et al. (2006)

reported on the utilization of transcriptome and metabolome data in the elucidation of the terpenoid pathway in periwinkle cells (Rischer et al. 2006). This study showed a strong link between the gene-to-gene and gene-to-metabolite networks in the analysis of the terpenoid pathway. Through the utilization of a combinatorial approach, novel gene functions and pathways were predicted. Through this network, undescribed genes in terpernoid biosynthesis were identified and their function and pathways determined (Rischer et al. 2006). However while this network allows for determination between the gene and metabolite, a large portion of information on the relationship between genes, transcripts, proteome, and metabolome still remains elusive (Mounet et al. 2009).

16.1.6 The Bioinformatics Approach in Plant Secondary Metabolite Analysis

The advent of the omics era has resulted in a flood of omics data through various transcriptomic, proteomic, and metabolomic experiments. While having loads of data can be a good thing, this poses a challenge to data mining, extracting of useful information, deciphering the biological significance of the output, and generating the link from genome all the way to the metabolome.

In data mining, some of the commonly used methods would include multivariate statistical methods like PCA, PLS, HCA, SOM, BL-SOM, and Pearson correlation coefficient (Saito et al. 2006; Lindon et al. 2007). PCA is used in unsupervised data analysis and dimensional reduction where these are expressed variances of datasets expressed in PC factors. The conversion of matrix to PCs results in scores and loading matrices that generates plots that can be visualized (Jansen et al. 2005). Unlike PCA, PLS is used in supervised analysis when additional information on sample is available. Qualitative data is referred to as discriminatory analysis compared to quantitative (Saito et al. 2006). The PLS toolbox, MATLAB, and Pirouette can be used in conducting PCA and Pearson correlation coefficient analysis (Rischer et al. 2006). Further tools such as Cytoscape and Tom Sawyer could be used to generate network correlations (Rischer et al. 2006; Weckwerth 2008). Like PCA, HCA is another unsupervised method utilized in omics data analysis. The data generated by this tool is shown as a tree where two samples that are in close proximity on the tree are the most closely related (Yang et al. 2014).

SOM demonstrates the ability to identify related regions within the map. Nonetheless, for SOM to provide good information, it relies upon training on datasets. BL-SOM has the ability to overcome the drawbacks of SOM and generate reproducible maps. BL-SOM is a complex type of multivariate inquiry, which facilitates the characterization of transcripts and metabolites into cells on 2D lattices, where samples with similar or identical patterns are clustered together into the same region or space within the cells (Hirai et al. 2005; Lindon et al. 2007). Pattern recognition is

approached through artificial neural networks. Correlation analysis, cluster formation, and SOMs have been successfully analyzed through the utilization of GeneMaths software (Mercke et al. 2004).

Metabolomic resources are an important component in analyzing secondary metabolites. One good web resource that is able to analyze, visualize, and annotate large metabolomics data is MetaboAnalyst (Xia et al. 2012). Further both KaPPA-View and PRIMe are two online resources that are useful in the event there is a need to integrate transcript to the metabolite information in an analysis (Tokimatsu et al. 2005; Akiyama et al. 2008). Integrated information from genomics to proteomics may be analyzed through Babelomics (Medina et al. 2010). A large number of resources have been developed specifically for certain plants, especially for the model plant *Arabidopsis thaliana*. Some examples of these resources are tools such as PathMAPA (Pan et al. 2003), Metnet (Wurtele et al. 2003), MetaCyc (Zhang et al. 2005), AraCyc (Zhang et al. 2005), Pathway Processor (Beltrame et al. 2013), and MapMan (Thimm et al. 2004). Other metabolomics resources available online are MeRy-B and MetabolomeExpress (Akiyama et al. 2008; Carroll et al. 2010; Ferry-Dumazet et al. 2011). Lists of tools and databases available for in silico analysis are provided in Tables 16.1 and 16.2, respectively.

16.1.7 In Silico Systems for Emerging Plant Systems

Most bioinformatics tools were initially designed and trained for the *Arabidopsis thaliana* genome analysis. Since then many more model and non-model crops have been sequenced and their databases deposited for use by the scientific community. Tools have also been developed specifically for some of these crops. The following paragraphs will highlight some of these main plant groups.

16.1.7.1 Graminaceae

This group contains most grasses and food crops which include rice, maize, wheat, and barley. This is also the group with the largest number of sequenced genomes, beginning with the rice genome which acts as a model plant genome for all proceeding studies in this group (Somerville et al. 2010; Lobell et al. 2011). Following the completion of the Rice Genome Project, whole-genome sequences have been obtained for barley, wheat, maize, sorghum, and *Brachypodium* (Paterson et al. 2009; Schnable et al. 2009). The genome and EST database that are available for all the abovementioned crops have made it easier to identify new novel genes and to predict gene function in these crops through correlations between gene-to-gene and gene-to-metabolite (Mochida et al. 2003, 2006, 2008; Ogihara et al. 2003; Zhang et al. 2004; Kawaura et al. 2006; Sato et al. 2009). Further transcriptome

Table 16.1 Tools for in silico analysis of plant metabolites

Tools	Type of data	Organism	Application	Website
MetaboAnalyst	Metabolomic (NMR/MS and LC/GC-MS)	All plants	Metabolomic, multivariate statistical analysis and annotations	http://www.metaboanalyst.ca
ASCA	Metabolomic	All plants	Complex multivariate analysis	http://www.bdagroup.nl/
KaPPA-View	Transcriptomic and metabolomic	*Arabidopsis*, rice, *Lotus japonicus*, and tomato	Gene -to-gene and metabolite-to-metabolite predictions	https://omictools.com/kappa-view-tool
Babelomics	Transcriptomic, proteomics, and genomics	All plants	Functional profiling of genomic data	http://www.babelomics.org
MapMan	Transcriptomic and metabolomic	All plants, applied successfully in *Arabidopsis* and maize	Metabolic pathways and processes	http://gabi.rzpd.de/projects/MapMan/
Metnet	mRNA, protein, and metabolite profiling	*Arabidopsis*, soybean	Visualize statistically and analyze metabolic and regulatory mapping	http://metnetonline.org/
PathMAPA	Microarray	*Arabidopsis*	Metabolic pathway analysis	http://bioinformatics.med.yale.edu/pathmapa.htm
PRIMe	Metabolomics including NMR spectroscopy, GC/MS, LC/MS, and CE/MS	General plants; applied successful in *Arabidopsis*	Integration of gene expression and metabolite accumulation	http://prime.psc.riken.jp
plantiSMASH	A versatile online analysis platform that automates the identification of plant biosynthetic gene clusters	All plants	Identification of biosynthetic gene clusters and comparative genomics of each cluster	https://academic.oup.com/nar/article/45/W1/W55/3769247

Adapted and modified from Yang et al. (2014)

analyses of Graminaceae have been successfully conducted using online resources such as Genevestigator, RiceArrayNet, and OryzaExpress (Lee et al. 2009; Hamada et al. 2011). Phosphoproteome Database and OryzaPG-DB that contains information on rice proteins (Nakagami et al. 2010, 2011; Helmy et al. 2011) have been utilized in the proteome analysis of the species within this group. Collectively the tools and databases mentioned above are supportive of integrated omics studies on secondary metabolites in plant systems (Mochida and Shinozaki 2011).

Table 16.2 Databases available for plant omics analysis. Suitable for secondary metabolite analysis

Name	Databases	Organism	Application	Website
MetaCyc	Pathway and genome databases	Multiple organisms	Prediction of metabolic pathways	http://metacyc.org
AraCyc	Biochemical pathway	*Arabidopsis thaliana*	Represents *Arabidopsis* metabolism	http://www.arabidopsis.org/tools/aracyc
MediCyc	Genomics and metabolic pathway	*M. truncatula*	To predict the metabolic composition for *M. truncatula*	http://medicyc.noble.org/
MetaCrop	Metabolic pathways in crop plants	Crop plants	Allows automatic export of information for the creation of metabolic models	http://metacrop.ipk-gatersleben.de/apex/f?p=269:111:
Golm Metabolome Database (GMD)	Contains reference mass spectra and metabolite profiling	Multiorganisms	Facilitates the search for and dissemination of reference mass spectra from metabolites quantified using GC/MS	http://gmd.mpimp-golm.mpg.de/
KNApSAcK	A comprehensive plant species-metabolite relationship database	All plants	Describes the relationships between species and their metabolites and would be useful for metabolomics research	http://kanaya.aist-nara.ac.jp/KNApSAcK/

16.1.7.2 Solanaceae

The Solanaceae species includes crops such as tomatoes and potatoes. These crops mostly have small genomes and show a high level of synteny. The International Tomato Annotation Group completed and provided the annotation for the sequences of the tomato genome. The tomato transcriptome data was also made available through NCBI GEO, and consequently advances were made in metabolomics where metabolome profiles were also made public (Do et al. 2010; Enfissi et al. 2010; Schilmiller et al. 2010; Tieman et al. 2010) in platforms such as Plant MetGenMAP, MotoDB, KOMICS, and KaPPA-View4 SOL (Iijima et al. 2008; Joung et al. 2009; Sakurai et al. 2011). The other Solanaceae that were sequenced are potatoes and peppers. The potato clone sequence is made available at http://www.potatogenome.net/index.php/Main_Page (Xu et al. 2011), while a collection of annotated pepper ESTs at http://genepool.kribb.re.kr/pepper/ (Kim et al. 2008). While there are these individual databases for the Solanaceae, an integrated

collection of all these ESTs can be accessed at the SolEST database (D'Agostino et al. 2009). The Sol Genomics Network is an online resource for Solanaceae research that provides a comprehensive, in-depth information on genomic resources for Solanaceae species and their close relatives (Bombarely et al. 2011; Mochida and Shinozaki 2011).

16.1.7.3 Leguminosae

The legumes are a nitrogen-fixing, symbiotic group of crops. These crops act as an excellent model to study nitrogen fixation and symbiosis. As such post sequencing of *L. japonicus*, the TILLING resource was utilized to study genes associated with symbiosis (Perry et al. 2009), while the proteome was studied in the seed and pod as part of studying their efficiency in nitrogen fixation and storage (Nautrup-Pedersen et al. 2010). *Medicago truncatula* Genome Project in JCVI/TIGR provides information on transcriptome, mutations, reverse genetics, and pseudomolecules (Young and Udvardi 2009). The *Medicago truncatula* HapMap Project acts as a reference sequence and provides the GWAS resource. In addition to the references above, there are also the MIRMED Database and the Medicago PhosphoProtein Database which act as a source for miRNAs and proteins produced by this plant genome, respectively (Grimsrud et al. 2010a, b). Similarly an annotated genome (Glyma1.0) and a transcriptome resource for soybean have been developed using NGS and RNA-seq platforms (Libault et al. 2010). Further an informative database on soybean is provided in the online portal SoyBase which integrates resources from various soybean research (Grant et al. 2010; Mochida and Shinozaki 2011).

16.1.8 Concluding Remarks and Future Prospects

For decades plant secondary metabolites have been widely sourced by the food, energy, and pharmaceutical industries. While plants are a good source of secondary metabolites, its concentration in plants remains low. Secondary metabolite synthesis is influenced by plant developmental stages and environment (Alagna et al. 2009; Oksman-Caldentey and Saito 2005). Due to the economic potential of this compound, interest has continued in sourcing these compounds since the tracer technology (Saito et al. 2008). The limited quantity and the growing need for these compounds have led to overharvesting of plants for various human needs (Alagna et al. 2009). This therefore has directed the course of studies to metabolic engineering. While some studies in metabolic engineering have resulted in successfully engineered plants, a large number of these attempts have not yielded positive results (Naoumkina et al. 2010; Capell and Christou 2004; Wu and Chappell 2008; Dudareva and Pichersky 2008). While there have been attempts at overexpressing some of these compounds to increase plant defense, the genetic modifications of genes or pathways have resulted in

unexpected products or perhaps even in no change in level of secondary metabolite within the plant. In certain cases these overexpressions have led to inhibition of product (Naoumkina et al. 2010; Yang et al. 2014).

Omics tools, especially the integration of the omics platforms, open new avenues to decipher secondary metabolism and direct metabolic engineering (Oksman-Caldentey and Saito 2005; Trethewey 2001; Arnaud et al. 2007; Kizer et al. 2008). In the past two decades, various online tools have been developed to conduct integrated omics analysis. Through the utilization of these online resources for functional analysis, many new novel genes have been identified. Through these resources, gene function was comprehended at a systemic level, and the genes have been incorporated into novel pathways for a better understanding of the process involved in the biosynthesis of metabolites. The information derived from these in silico analyses will provide beneficial information that may be utilized to set the parameters for metabolic engineering. The major constraint in our ability to conduct metabolic engineering is our lack of knowledge of the networks involved in synthesis of secondary metabolites (Yang et al. 2014).

While in silico analysis definitely provides information on which we may design our experiments and our engineering, we are still a long way from having systems that provide us with definite information on the metabolic systems and models. We still need to overcome the shortcoming in metabolic modeling before it can be incorporated into routinely being incorporated as part of crop systems biology analysis (Baldazzi et al. 2012). One possible solution to this may be the development of a multiscale model that may incorporate the information from various spatial and temporal resources which includes the biological, biochemical, and other processes (Baldazzi et al. 2012; Grieneisen et al. 2012; Vernoux et al. 2011; Walpole et al. 2013). Therefore while the omics technologies and tools advance with time to assist with the study of secondary metabolites, there is much that is known and remains obscure. The bioinformatics sector will have to continuously evolve and improve their resources so that they are better and faster and best represent the biological process in silico, hence making the predictions more accurate.

References

Akiyama K, Chikayama E, Yuasa H et al (2008) PRIMe: a web site that assembles tools for metabolomics and transcriptomics. In Silico Biol 8:339–345. PMID: 19032166

Alagna F (2013) Innovative transcriptomics approaches for large scale identification of genes involved in plant secondary metabolism. J Plant Biochem Physiol 1:e107. https://doi.org/10.4172/2329-9029.1000e107

Alagna F, D'Agostino N, Torchia L et al (2009) Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. BMC Genomics 10:399. https://doi.org/10.1186/1471-2164-10-399

Arnaud B, Elio S, Robert DH (2007) Metabolic engineering of flavonoids in tomato (*Solanum lycopersicum*): the potential for metabolomics. Metabolomics 3:399–412. https://doi.org/10.1007/s11306-007-0074-2

Baldazzi V, Bertin N, de Jong H et al (2012) Towards multiscale plant models: integrating cellular networks. Trends Plant Sci 17:728–736. https://doi.org/10.1016/j.tplants.2012.06.012

Beckers V, Dersch LM, Lotz K et al (2016) In silico metabolic network analysis of Arabidopsis leaves. BMC Syst Biol 10:102. https://doi.org/10.1186/s12918-016-0347-3

Beltrame L, Bianco L, Fontana P et al (2013) Pathway processor 2.0: a web resource for pathway-based analysis of high-throughput data. Bioinformatics 29:1825–1826. https://doi.org/10.1093/bioinformatics/btt292

Bombarely A, Menda N, Tecle IY et al (2011) The sol genomics network (solgenomics.net): growing tomatoes using Perl. Nucleic Acids Res 39:D1149–D1155. https://doi.org/10.1093/nar/gkq866

Broun P (2004) Transcription factors as tools for metabolic engineering in plants. Curr Opin Plant Biol 7:202–209. https://doi.org/10.1016/j.pbi.2004.01.013

Broun P (2005) Transcriptional control of flavonoid biosynthesis: a complex network of conserved regulators involved in multiple aspects of differentiation in *Arabidopsis*. Curr Opin Plant Biol. 2005 8:272–279. https://doi.org/10.1016/j.pbi.2005.03.006

Capell T, Christou P (2004) Progress in plant metabolic engineering. Curr Opin Biotechnol 15:148–154. https://doi.org/10.1016/j.copbio.2004.01.009

Carelli M, Biazzi E, Panara F, Tava A et al (2011) *Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. Plant Cell 23:3070–3081. https://doi.org/10.1105/tpc.111.087312

Carroll AJ, Badger MR, Harvey Millar A (2010) The MetabolomeExpress project: enabling web-based processing, analysis and transparent dissemination of GC/MS metabolomics datasets. BMC Bioinforma 11:376. https://doi.org/10.1186/1471-2105-11-376

Cusido RM, Onrubia M, Sabater-Jara AB et al (2014) A rational approach to improving the biotechnological production of taxanes in plant cell cultures of Taxus spp. Biotechnol Adv 32 (6):1157–1167. https://doi.org/10.1016/j.biotechadv.2014.03.002

D'Agostino N, Traini A, Frusciante L, Chiusano ML (2009) SolEST database: a 'one-stop shop' approach to the study of Solanaceae transcriptomes. BMC Plant Biol 9:142. https://doi.org/10.1186/1471-2229-9-142

de Oliveira Dal'Molin CG, Quek LE, Palfreyman RW et al (2010) AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. Plant Physiol 152(2):579–589. doi.org/10.1104/pp.109.148817

Dersch LM, Beckers V, Wittmann C (2016) Green pathways: metabolic network analysis of plant systems. Metab Eng 34:1–24. https://doi.org/10.1016/j.ymben.2015.12.001

Do PT, Prudent M, Sulpice R et al (2010) The influence of fruit load on the tomato pericarp metabolome in a *Solanum chmielewskii* introgression line population. Plant Physiol 154:1128–1142. https://doi.org/10.1104/pp.110.163030

Dudareva N, Pichersky E (2008) Metabolic engineering of plant volatiles. Curr Opin Biotechnol 2008(19):181–189. https://doi.org/10.1016/j.copbio.2008.02.011

Dyer JM, Stymne S, Green AG et al (2008) High-value oils from plants. Plant J 54:640–655. https://doi.org/10.1111/j.1365-313X.2008.03430.x

Enfissi EM, Barneche F, Ahmed I et al (2010) Integrative transcript and metabolite analysis of nutritionally enhanced DE-ETIOLATED1 down-regulated tomato fruit. Plant Cell 22:1190–1215. https://doi.org/10.1105/tpc.110.073866

Ferry-Dumazet H, Gil L, Deborde C et al (2011) MeRy-B: a web knowledgebase for the storage, visualization, analysis and annotation of plant NMR metabolomic profiles. BMC Plant Biol 11:104. https://doi.org/10.1186/1471-2229-11-104

Fiehn O (2001) Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. Comp Funct Genomics 2:155–168. https://doi.org/10.1002/cfg.82

Fiehn O, Weckwerth W (2003) Deciphering metabolic networks. Eur J Biochem 270:579–588. PMID:12581198

Fiehn O, Kopka J, Dormann P et al (2000) Metabolite profiling for plant functional genomics. Nat Biotechnol 18:1157–1161. https://doi.org/10.1038/81137

Field B, Osbourn AE (2008) Metabolic diversification – independent assembly of operon-like gene clusters in different plants. Science 320:543–547. https://doi.org/10.1126/science.1154990
Field B, Fiston-Lavier AS, Kemen A, Geisler K, Quesneville H et al (2011) Formation of plant metabolic gene clusters within dynamic chromosomal regions. Proc Natl Acad Sci U S A 108:16116–16121. https://doi.org/10.1073/pnas.1109273108
Forester SC, Waterhouse AL (2009) Metabolites are key to understanding health effects of wine polyphenolics. J Nutr 139:1824S–1831S. https://doi.org/10.3945/jn.109.107664
Forster J, Famili I, Fu P et al (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Res 13:244–253. https://doi.org/10.1101/gr.234503
Grafahrend-Belau E, Schreiber F, Koschützki D et al (2009) Flux balance analysis of barley seeds: a computational approach to study systemic properties of central metabolism. Plant Physiol 149 (1):585–598. https://doi.org/10.1104/pp.108.129635
Grant D, Nelson RT, Cannon SB et al (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Res 38:D843–D846. https://doi.org/10.1093/nar/gkp798
Grieneisen VA, Scheres B, Hogeweg P et al (2012) Morphogengineering roots: comparing mechanisms of morphogen gradient formation. BMC Syst Biol 6:37. https://doi.org/10.1186/1752-0509-6-37
Grimsrud PA, Swaney DL, Wenger CD et al (2010a) Phosphoproteomics for the masses. ACS Chem Biol 5:105–119. 10.1021/cb900277e. https://doi.org/10.1021/cb900277e
Grimsrud PA, den Os D, Wenger CD et al (2010b) Large-scale phosphoprotein analysis in *Medicago truncatula* roots provides insight into in vivo kinase activity in legumes. Plant Physiol 152:19–28. 10.1104/pp.109.149625. https://doi.org/10.1104/pp.109.149625
Hamada K, Hongo K, Suwabe K et al (2011) OryzaExpress: an integrated database of gene expression networks and omics annotations in rice. Plant Cell Physiol 52:220–229. https://doi.org/10.1093/pcp/pcq195
Hammami R, Ben-Hamida J, Vergoten G et al (2009) Phytamp: a database dedicated to antimicrobial plant peptides. Nucl Acids Res 37:D963–D968. https://doi.org/10.1093/nar/gkn655
Hartmann T (2007) From waste products to ecochemicals: fifty years research of plant secondary metabolism. Phytochemistry 68:2831–2846. https://doi.org/10.1016/j.phytochem.2007.09.017
Helmy M, Tomita M, Ishihama Y (2011) OryzaPG-DB: rice proteome database based on shotgun proteogenomics. BMC Plant Biol 11:63. https://doi.org/10.1186/1471-2229-11-63
Hirai MY, Klein M, Fujikawa Y et al (2005) Elucidation of gene-to-gene and metabolite-to-gene networks in Arabidopsis by integration of metabolomics and transcriptomics. J Biol Chem 280:25590–25595. https://doi.org/10.1074/jbc.M502332200
Hostettman K, Terreaux C (2000) Search for new lead compounds from higher plants. Chimia 54:652–657. ISSN 0009-4293
Hwang KS, Kim HU, Charusanti P et al (2014) Systems biology and biotechnology of Streptomyces species for the production of secondary metabolites. Biotechnol Adv 32(2):255–268. https://doi.org/10.1016/j.biotechadv.2013.10.008
Iijima Y, Nakamura Y, Ogata Y et al (2008) Metabolite annotations based on the integration of mass spectral information. Plant J 54:949–962. https://doi.org/10.1111/j.1365-313X.2008.03434.x
Jansen JJ, Hoefsloot HCJ, van der Greef J et al (2005) Multilevel component analysis of time-resolved metabolic fingerprinting data. Anal Chim Acta 530:173–183. https://doi.org/10.1016/j.aca.2004.09.074
Joung JG, Corbett AM, Fellman SM et al (2009) Plant MetGenMAP: an integrative analysis system for plant systems biology. Plant Physiol 151:1758–1768. https://doi.org/10.1104/pp.109.145169
Junker BH (2014) Flux analysis in plant metabolic networks: increasing throughput and coverage. Curr Opin Biotechnol 26:183–188. https://doi.org/10.1016/j.copbio.2014.01.016
Kawaura K, Mochida K, Yamazaki Y et al (2006) Transcriptome analysis of salinity stress responses in common wheat using a 22k oligo-DNA microarray. Funct Integr Genomics 6:132–142. https://doi.org/10.1007/s10142-005-0010-3

Kim HJ, Baek KH, Lee SW et al (2008) Pepper EST database: comprehensive in silico tool for analyzing the chili pepper (*Capsicum annuum*) transcriptome. BMC Plant Biol 8:101. https://doi.org/10.1186/1471-2229-8-101

Kim B, Park H, Na D, Lee SY (2014) Metabolic engineering of *Escherichia coli* for the production of phenol from glucose. Biotechnol J 9(5):621–629. https://doi.org/10.1002/biot.201300263

Kizer L, Pitera DJ, Pfleger BF et al (2008) Application of functional genomics to pathway optimization for increased isoprenoid production. Appl Environ Microbiol 74:3229–3241. https://doi.org/10.1128/AEM.02750-07

Lee TH, Kim YK, Pham TT et al (2009) RiceArrayNet: a database for correlating gene expression from transcriptome profiling, and its application to the analysis of coexpressed genes in rice. Plant Physiol 151:16–33. https://doi.org/10.1104/pp.109.139030

Li GL, Kollner TG, Yin Y et al (2012) Nonseed plant *Selaginella moellendorffii* has both seed plant and microbial types of terpene synthases. Proc Natl Acad Sci U S A 109:14711–14715. https://doi.org/10.1073/pnas.1204300109

Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ, Franklin LD et al (2010) An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. Plant J 63:86–99. https://doi.org/10.1111/j.1365-313X.2010.04222.x

Lichtenthaler HK (2000) Non-mevalonate isoprenoid biosynthesis: enzymes, genes and inhibitors. Biochem Soc Trans 28:785–789. PMID:11171208

Lindon JC, Nicholson JK, Holmes E (2007) The handbook of metabonomics and metabolomics. Elsevier, Amsterdam. ISBN:9780080468006

Lobell DB, Schlenker W, Costa-Roberts J (2011) Climate trends and global crop production since 1980. Science 333:616–620. https://doi.org/10.1126/science.1204531

Lotz K, Hartmann A, Grafahrend-Belau E, Schreiber F, Junker BH (2014) Elementary flux modes, flux balance analysis, and their application to plant metabolism. Methods Mol Biol 1083:231–252. https://doi.org/10.1007/978-1-62703-661-0_14

Manners JM (2007) Hidden weapons of microbial destruction in plant genomes. Genome Biol 8:225. https://doi.org/10.1186/gb-2007-8-9-225

Medina I, Carbonell J, Pulido L et al (2010) Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. Nucleic Acids Res 38:210–213. https://doi.org/10.1093/nar/gkq388

Mercke P, Kappers IF, Verstappen FWA et al (2004) Combined transcript and metabolite analysis reveals genes involved in spider mite induced volatile formation in cucumber plants. Plant Physiol 135:2012–2024. https://doi.org/10.1104/pp.104.048116

Metlen KL, Aschehoug ET, Callaway RM (2009) Plant behavioural ecology: dynamic plasticity in secondary metabolites. Plant Cell Environ 32:641–653. https://doi.org/10.1111/j.1365-3040.2008.01910.x

Metzker ML (2010) Sequencing technologies—the next generation. Nat Rev Genet 11:31–46. https://doi.org/10.1038/nrg2626

Michael TP, Jackson S (2013) The first 50 plant genomes. Plant Genome 6. https://doi.org/10.3835/plantgenome2013.03.0001in

Mochida K, Shinozaki K (2011) Advances in omics and bioinformatics tools for systems analyses of plant functions. Plant Cell Physiol 52(12):2017–2038. https://doi.org/10.1093/pcp/pcr153

Mochida K, Yamazaki Y, Ogihara Y (2003) Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. Mol Gen Genomics 270:371–377. https://doi.org/10.1007/s00438-003-0939-7

Mochida K, Kawaura K, Shimosaka E et al (2006) Tissue expression map of a large number of expressed sequence tags and its application to in silico screening of stress response genes in common wheat. Mol Gen Genomics 276:304–312. https://doi.org/10.1007/s00438-006-0120-1

Mochida K, Saisho D, Yoshida T et al (2008) TriMEDB: a database to integrate transcribed markers and facilitate genetic studies of the tribe Triticeae. BMC Plant Biol 8:72. https://doi.org/10.1186/1471-2229-8-72

Mounet F, Moing A, Garcia V et al (2009) Gene and metabolite regulatory network analysis of early developing fruit tissues highlights new candidate genes for the control of tomato fruit composition and development. Plant Physiol 149:1505–1528. https://doi.org/10.1104/pp.108.133967

Nakagami H, Sugiyama N, Mochida K et al (2010) Large-scale comparative phosphoproteomics identifies conserved phosphorylation sites in plants. Plant Physiol 153:1161–1174. https://doi.org/10.1104/pp.110.157347

Nakagami H, Sugiyama N, Ishihama Y et al (2011) Shotguns in the front line: phosphoproteomics in plants. Plant Cell Physiol 53(1):118–124. https://doi.org/10.1093/pcp/pcr148

Naoumkina MA, Modolo LV, Huhman DV et al (2010) Genomic and coexpression analyses predict multiple genes involved in triterpene saponin biosynthesis in *Medicago truncatula*. Plant Cell 22:850–866. https://doi.org/10.1105/tpc.109.073270

Nautrup-Pedersen G, Dam S, Laursen BS et al (2010) Proteome analysis of pod and seed development in the model legume *Lotus japonicus*. J Proteome Res 9:5715–5726. https://doi.org/10.1021/pr100511u

Nawrot R, Barylski J, Nowicki G et al (2014) Plant antimicrobial peptides. Folia Microbiol (Praha) 59(3):181–196. https://doi.org/10.1007/s12223-013-0280-4

Ogihara Y, Mochida K, Nemoto Y et al (2003) Correlated clustering and virtual display of gene expression patterns in the wheat life cycle by large-scale statistical analyses of expressed sequence tags. Plant J 33:1001–1011. PMID:12631325

Oksman-Caldentey KM, Saito K (2005) Integrating genomics and metabolomics for engineering plant metabolic pathways. Curr Opin Biotechnol 16:174–179. https://doi.org/10.1016/j.copbio.2005.02.007

Pan D, Sun N, Cheung KH et al (2003) PathMAPA: a tool for displaying gene expression and performing statistical tests on metabolic pathways at multiple levels for Arabidopsis. BMC Bioinforma 4:56. https://doi.org/10.1186/1471-2105-4-56

Paterson AH, Bowers JE, Bruggmann R et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556. https://doi.org/10.1038/nature07723

Pereira DM, Valentao P, Correia-da-Silva G et al (2012) Plant secondary metabolites in cancer chemotherapy: where are we? Curr Pharm Biotechnol 13:632–650. PMID:22122478

Perry J, Brachmann A, Welham T et al (2009) TILLING in Lotus japonicus identified large allelic series for symbiosis genes and revealed a bias in functionally defective ethyl methanesulfonate alleles toward glycine replacements. Plant Physiol 151:1281–1291. https://doi.org/10.1104/pp.109.142190

Pestana-Calsa MC, Ribeiro IL, Calsa T Jr (2010) Bioinformatics-coupled molecular approaches for unravelling potential antimicrobial peptides coding genes in Brazilian native and crop plant species. Curr Protein Pept Sci 11:199–209. PMID:20088767

Poblete-Castro I, Binger D, Rodrigues A et al (2013) In-silico-driven metabolic engineering of *Pseudomonas putida* for enhanced production of poly-hydroxyalkanoates. Metab Eng 15:113–123. https://doi.org/10.1016/j.ymben.2012.10.004

Rischer H, Oresic M, Seppanen-Laakso T et al (2006) Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* cells. Proc Natl Acad Sci U S A 103:5614–5619. https://doi.org/10.1073/pnas.0601027103

Roberts SC (2007) Production and engineering of terpenoids in plant cell culture. Nat Chem Biol 3:387–395. https://doi.org/10.1038/nchembio.2007.8

Rohmer M (1999) The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. Nat Prod Rep 16:565–574. PMID:10584331

Saha R, Suthers PF, Maranas CD (2011) *Zea mays* iRS1563: a comprehensive genome-scale metabolic reconstruction of maize metabolism. PLoS One 6(7):e21784. https://doi.org/10.1371/journal.pone.0021784

Saito K, Dixon RA, Willmitzer L (2006) In: Saito K, Dixon RA, Willmitzer L (eds) Plant metabolomics. Springer-Verlag, Berlin, pp 111–113. https://doi.org/10.1007/3-540-29782-0

Saito K, Hirai MY, Yonekura-Sakakibara K (2008) Decoding genes with coexpression networks and metabolomics – 'Majority report by precogs'. Trends Plant Sci 13:36–43. https://doi.org/10.1016/j.tplants.2007.10.006

Sakurai N, Ara T, Ogata Y et al (2011) KaPPA-View4: a metabolic pathway database for representation and analysis of correlation networks of gene co-expression and metabolite co-accumulation and omics data. Nucleic Acids Res 39:D677–D684. https://doi.org/10.1093/nar/gkq989

Sato K, Nankaku N, Takeda K (2009) A high-density transcript linkage map of barley derived from a single population. Heredity 103:110–117. https://doi.org/10.1038/hdy.2009.57

Schatz MC, Witkowski J, McCombie WR (2012) Current challenges in de novo plant genome sequencing and assembly. Genome Biol 13:243. https://doi.org/10.1186/gb4015

Schilmiller A, Shi F, Kim J et al (2010) Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines. Plant J 62:391–403. https://doi.org/10.1111/j.1365-313X.2010.04154.x

Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115. https://doi.org/10.1126/science.1178534

Schoof H, Ernst R, Nazarov V et al (2004) MIPS Arabidopsis thaliana database (MAtDB): an integrated biological knowledge resource for plant genomics. Nucleic Acids Res 32:D373–D376. https://doi.org/10.1093/nar/gkh068

Shachar-Hill Y (2013) Metabolic network flux analysis for engineering plant systems. Curr Opin Biotechnol 24(2):247–255. https://doi.org/10.1016/j.copbio.2013.01.004

Silverstein KAT, Graham MA, Paape TD et al (2005) Genome organization of more than 300 defensin-like genes in Arabidopsis. Plant Physiol 138:600–610. doi.org/10.1104/pp.105.060079

Silverstein KAT, Moskal WA Jr, Wu HC et al (2007) Small cysteine-rich peptides resembling antimicrobial peptides have been underpredicted in plants. Plant J 51:262–280. https://doi.org/10.1111/j.1365-313X.2007.03136.x

Somerville C, Youngs H, Taylor C, Davis SC, Long SP (2010) Feedstocks for lignocellulosic biofuels. Science 329:790–792. https://doi.org/10.1126/science.1189268

Sweetlove LJ, Last RL, Fernie AR (2003) Predictive metabolic engineering: a goal for systems biology. Plant Physiol 132(2):420–425

Terras FR, Eggermont K, Kovaleva V et al (1995) Small cysteine-rich antifungal proteins from radish: their role in host defense. Plant Cell 7:573–588. doi.org/10.1105/tpc.7.5.573

Thimm O, Blasing O, Gibon Y et al (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J 37:914–939. PMID:14996223

Tieman D, Zeigler M, Schmelz E et al (2010) Functional analysis of a tomato salicylic acid methyl transferase and its role in synthesis of the flavor volatile methyl salicylate. Plant J 62:113–123. https://doi.org/10.1111/j.1365-313X.2010.04128.x

Tohge T, Nishiyama Y, Hirai MY et al (2005) Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. Plant J 42:218–235. https://doi.org/10.1111/j.1365-313X.2005.02371.x

Tokimatsu T, Sakurai N, Suzuki H et al (2005) KaPPAview: a web-based analysis tool for integration of transcript and metabolite data on plant metabolic pathway maps. Plant Physiol 138:1289–1300. https://doi.org/10.1104/pp.105.060525

Trethewey RN (2001) Gene discovery via metabolic profiling. Curr Opin Biotechnol 12:135–138. PMID:11287226

Ulrich-Merzenich G, Zeitler H, Jobst D et al (2007) Application of the "omic-" technologies in phytomedicine. Phytomedicine 14:70–82. https://doi.org/10.1016/j.phymed.2006.11.011

Urbanczyk-Wochniak E, Baxter C, Kolbe A et al (2005) Profiling of diurnal patterns of metabolite and transcript abundance in potato (*Solanum tuberosum*) leaves. Planta 221:891–903. https://doi.org/10.1007/s00425-005-1483-y

Vernoux T, Brunoud G, Farcot E et al (2011) The auxin signalling network translates dynamic input into robust patterning at the shoot apex. Mol Syst Biol 7:508. https://doi.org/10.1038/msb.2011.39

Verpoorte R, Memelink J (2002) Engineering secondary metabolite production in plants. Curr Opin Biotechnol 13:181–187. PMID:11950573

Walpole J, Papin JA, Peirce SM (2013) Multiscale computational models of complex biological systems. Annu Rev Biomed Eng 15(1):137–154. https://doi.org/10.1146/annurev-bioeng-071811-150104

Wang Y, Joshi T, Zhang XS et al (2006) Inferring gene regulatory networks from multiple microarray datasets. Bioinformatics 22:2413–2420. doi.org/10.1093/bioinformatics/btl396

Weckwerth W (2008) Integration of metabolomics and proteomics in molecular plant physiology – coping with the complexity by data-dimensionality reduction. Physiol Plant 132:176–189. https://doi.org/10.1111/j.1399-3054.2007.01011.x

Wu SQ, Chappell J (2008) Metabolic engineering of natural products in plants; tools of the trade and challenges for the future. Curr Opin Biotechnol 19:145–152. https://doi.org/10.1016/j.copbio.2008.02.007

Wurtele ES, Li J, Diao L et al (2003) MetNet: software to build and model the biogenetic lattice of Arabidopsis. Comp Funct Genomics 4:239–245. https://doi.org/10.1002/cfg.285

Xia J, Mandal R, Sinelnikov IV et al (2012) MetaboAnalyst 2.0 – a comprehensive server for metabolomic data analysis. Nucleic Acids Res 40:W127–W133. https://doi.org/10.1093/nar/gks374

Xu X, Pan S, Cheng S et al (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195. https://doi.org/10.1038/nature10158

Yan L, Kerr PS (2002) Genetically engineered crops: their potential use for improvement of human nutrition. Nutr Rev 60(5 Pt 1):135–141. PMID:12030276

Yang D, Du X, Yang Z et al (2014) Transcriptomics, proteomics, and metabolomics to reveal mechanisms underlying plant secondary metabolism. Eng Life Sci 00:1–11. https://doi.org/10.1002/elsc.201300075

Young ND, Udvardi M (2009) Translating *Medicago truncatula* genomics to crop legumes. Curr Opin Plant Biol 12:193–201. https://doi.org/10.1016/j.pbi.2008.11.005

Zhang H, Sreenivasulu N, Weschke W et al (2004) Large-scale analysis of the barley transcriptome based on expressed sequence tags. Plant J 40:276–290. https://doi.org/10.1111/j.1365-313X.2004.02209.x

Zhang P, Foerster H, Tissier CP et al (2005) MetaCyc and AraCyc. Metabolic pathway databases for plant research. Plant Physiol 138:27–37. https://doi.org/10.1104/pp.105.060376

Zhao J (2007) Nutraceuticals, nutritional therapy, phytonutrients, and phytotherapy for improvement of human health: a perspective on plant biotechnology application. Recent Pat Biotechnol 1:75–97. PMID:19075834