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

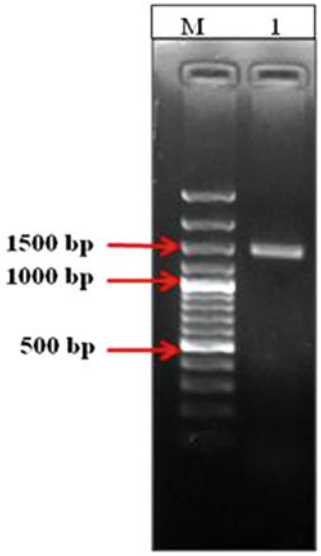
The nucleotide sequence of the 16S rRNA gene was submitted to NCBI GenBank, and accession number provided by the NCBI is KJ 184312. Bacterial culture was also deposited to Microbial Culture Collection (MCC), Pune, with accession number "MCC-2069" (Table 2)

Molecular Characterization of Plant Growth Promoting Bacterium and Phylogeny Analysis

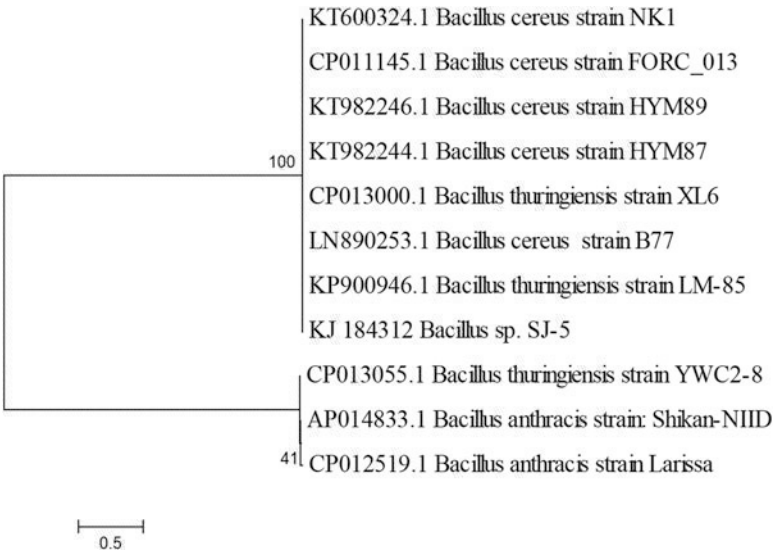
Molecular characterization of plant growth–promoting bacterium SJ-5 was done by amplifying 16S rRNA gene from genomic DNA by using universal 16S rRNA primers. On the agarose gel, a sharp band around 1.5 kb was observed. Sequence homology and phylogenetic analysis of the obtained sequence show similarity 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.

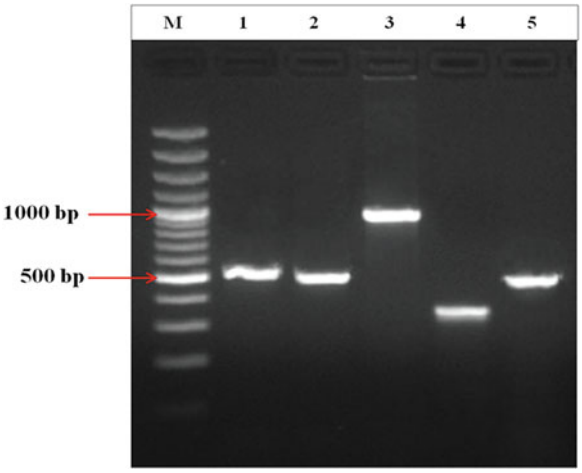


Evolutionary Relationships of Bacillus *sp. SJ-5*

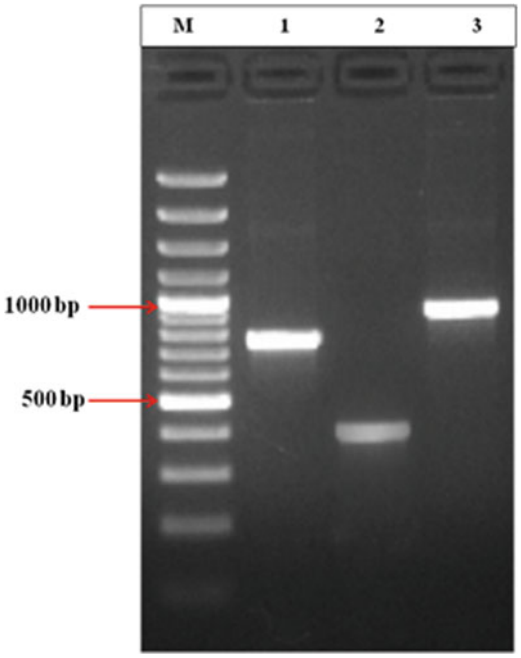
The evolutionary history was inferred using the neighbor-joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The evolutionary distances were computed using the maximum composite likelihood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

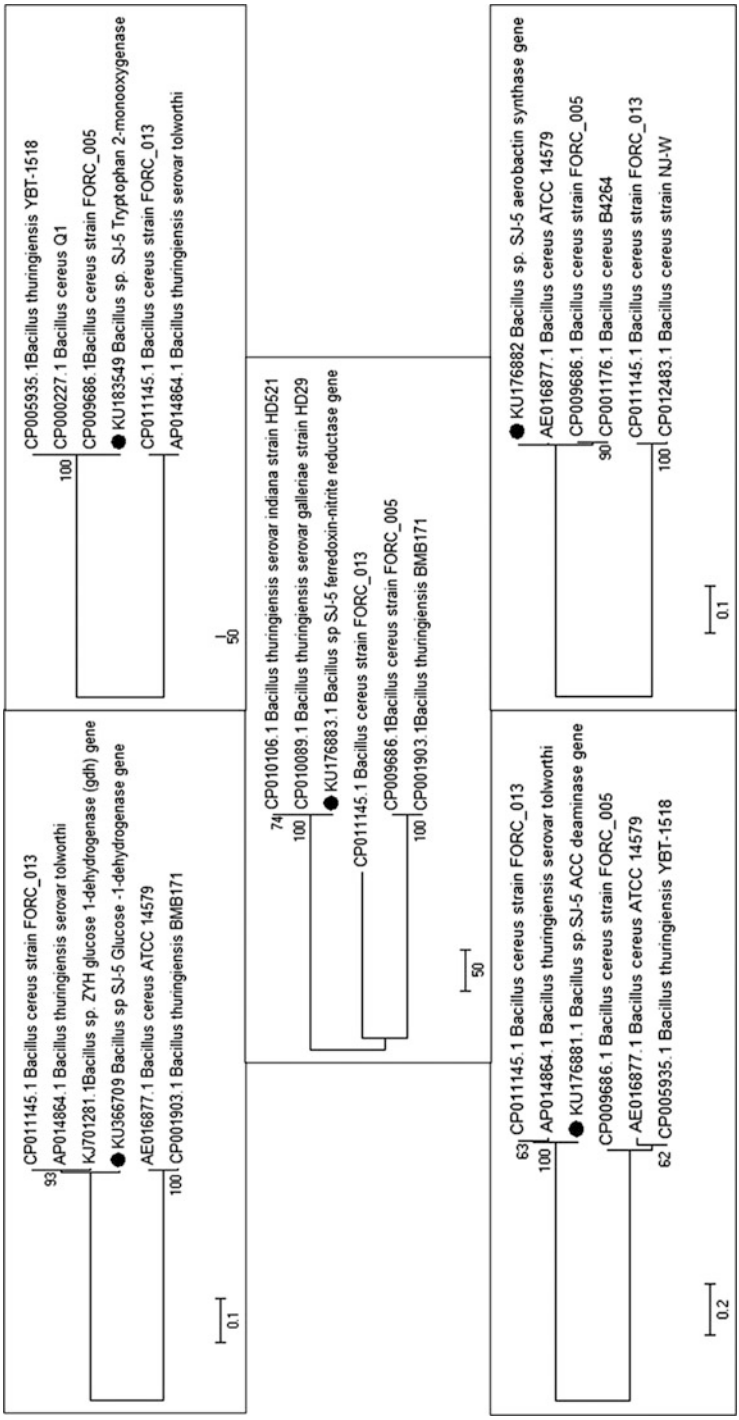


Fig. 1 Phylogenetic analysis of plant growth–promoting genes of *Bacillus* sp. SJ-5

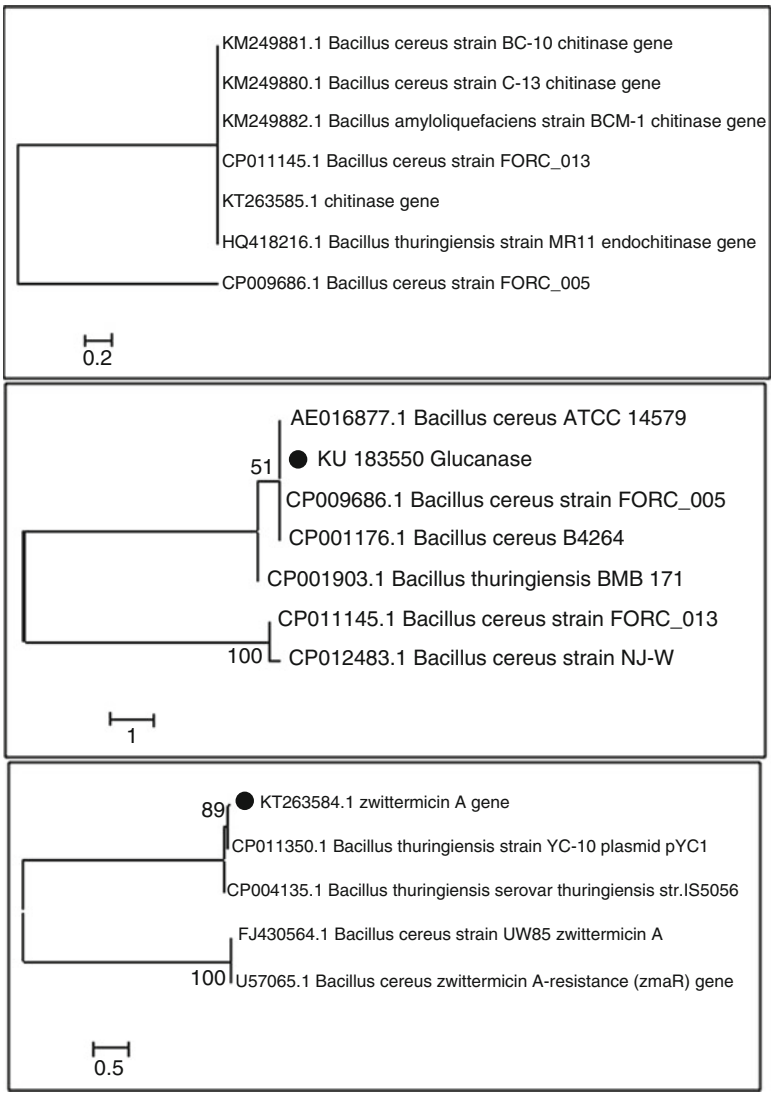


Fig. 2 Phylogenetic analysis of biocontrol genes of *Bacillus* sp. SJ-5

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

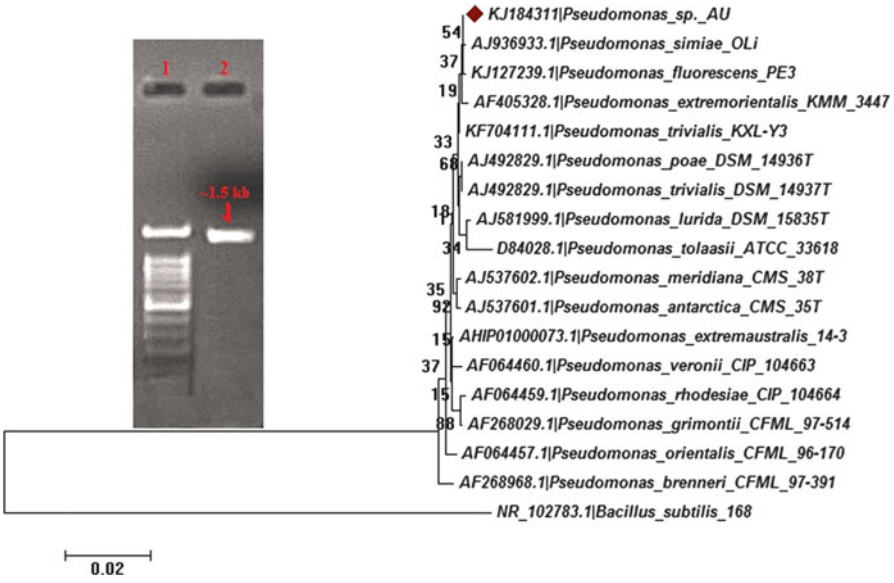


Fig. 3 Construction of phylogenetic tree based on 16S rRNA gene sequencing by neighbor-joining method

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

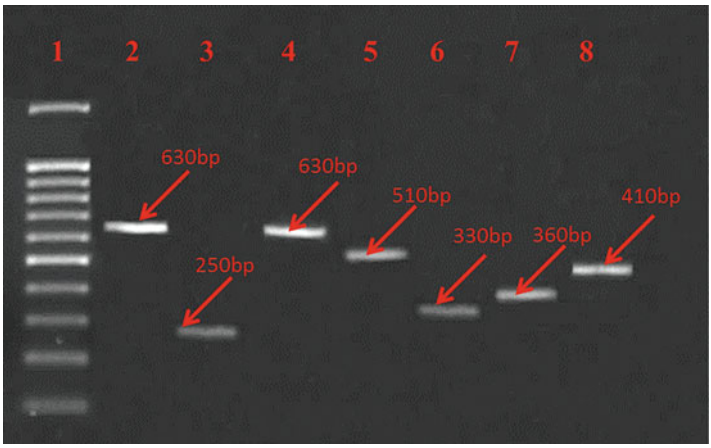


Fig. 4 PCR amplification of seven different PGP genes in AU bacterial isolate
Lane 1- DNA marker (100 bp), Lane 2- *g6pd*, Lane 3- *p5cr*, Lane 4- *nr,* Lane 5- *amy*, Lane 6- *IaaM*, Lane 7- *sid,* Lane 8- *acds*

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

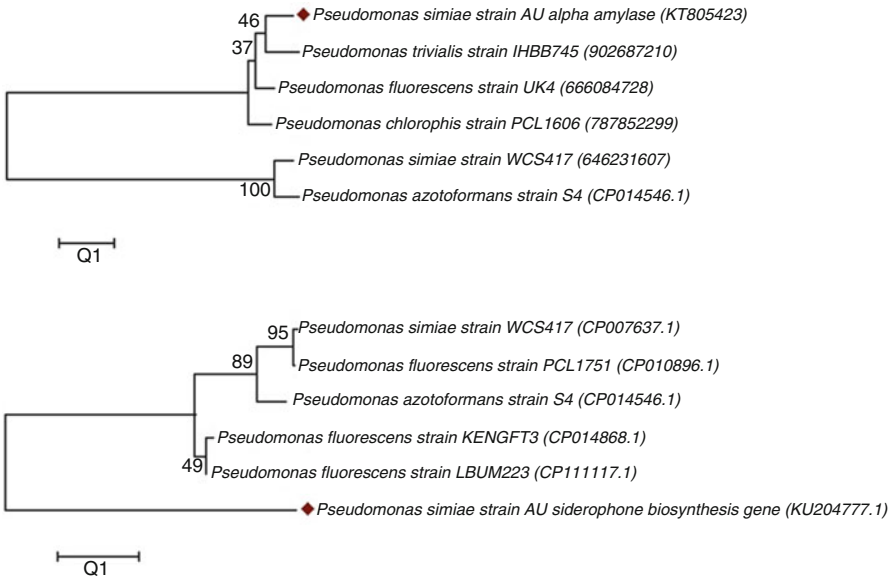


Fig. 5 Phylogenetic analysis of *acds* and *sid* gene of AU bacterial isolate

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

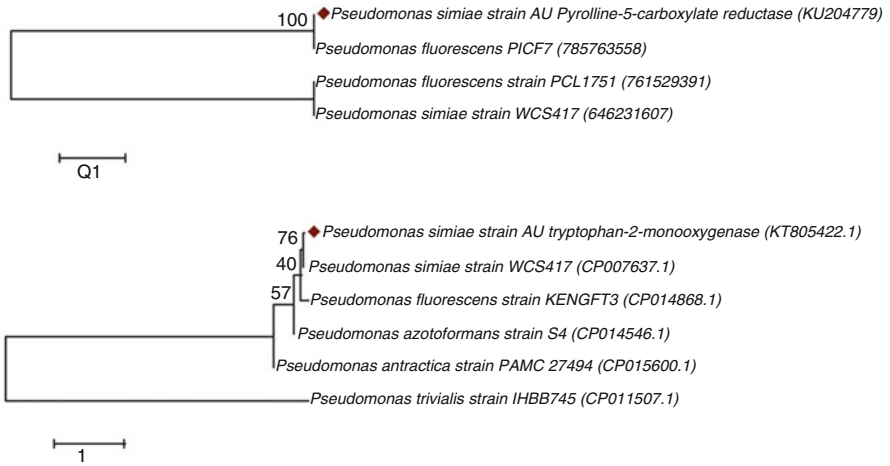


Fig. 6 Phylogenetic analysis of *p5cr* and *IaaM* isolated from AU bacterial isolate

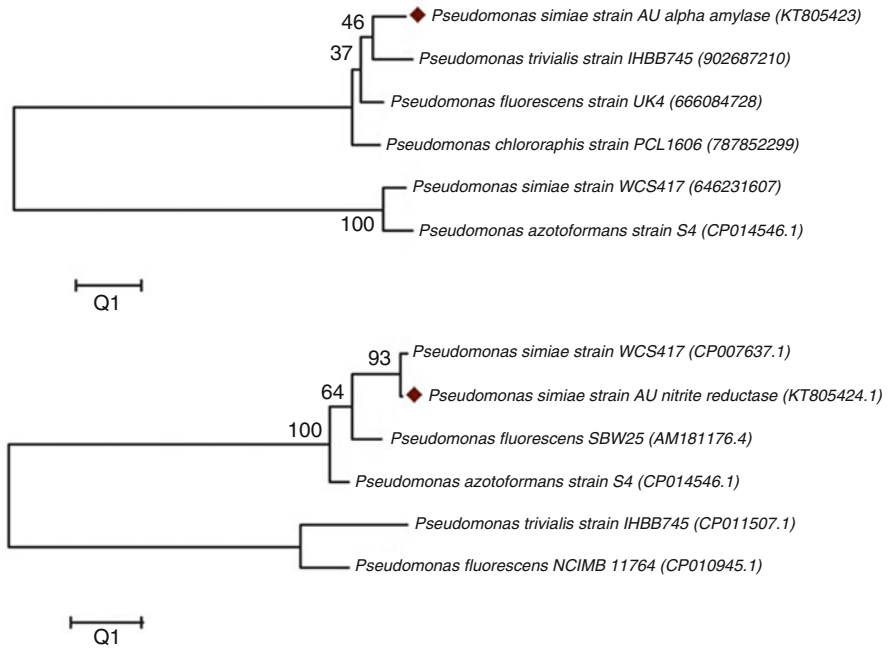


Fig. 7 Phylogenetic analysis of *amy* and *nr* isolated from AU bacterial isolate

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).

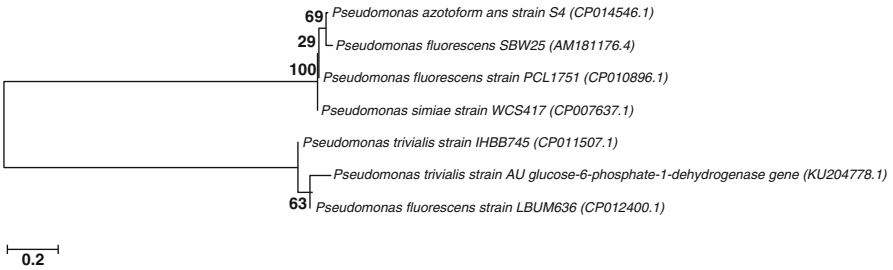


Fig. 8 Phylogenetic analysis of *g6pd* isolated from AU bacterial isolate

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

1 Genes to Metabolites and Metabolites to Genes Approaches to Predict Biosynthetic Pathways in Microbes for Natural Product Discovery 1
Dharmesh Harwani, Jyotsna Begani, and Jyoti Lakhani

2 Role of Functional Bacterial Phylum Proteobacteria in *Glycine max* Growth Promotion Under Abiotic Stress: A Glimpse on Case Study 17
Anukool Vaishnav, Amrita Kasotia, and Devendra Kumar Choudhary

3 Methanogenic Archaea in Paddy Agricultural Fields 51
Navnita Srivastva, Alpana Singh, and Suresh K. Dubey

4 In Silico Study of the Geminiviruses Infecting Ornamental Plants 69
Avinash Marwal, Megha Mishra, Rakesh Verma, Rajneesh Prajapat, and R. K. Gaur

5 In Silico Methods to Predict Disease-Resistance Candidate Genes in Plants 91
Jyoti Lakhani, Ajay Khuteta, Anupama Choudhary, and Dharmesh Harwani

6 In Silico Approach in Tracing Persistence and Survival of *Rhizobium* in Soil 107
Konineeka Sen and Chandan Sengupta

7 Bradyrhizobia-Mediated Drought Tolerance in Soybean and Mechanisms Involved 121
Abhishek Bharti, Richa Agnihotri, Hemant S. Maheshwari, Anil Prakash, and Mahaveer P. Sharma

8 In Silico Approach to Analyze the Biochemical Pathways of Bacterial Metabolite Synthesis 141
Tania, Mehendi Goyal, and Manoj Baranwal

9 In Silico Modulation Techniques for Upgrading Sustainability and Competitiveness in Agri-food Sector 161
Nitika Thakur

10 Bioinformatic Tools to Study the Soil Microorganisms: An In Silico Approach for Sustainable Agriculture 169
Pankaj Bhatt and Anupam Barh

11 Sustainable Agriculture: Role of Metagenomics and Metabolomics in Exploring the Soil Microbiota 183
Neeraj Gupta, Siddharth Vats, and Prachi Bhargava

12 In Silico Plant Growth Metabolites 201
Shradha Sharma and Pushpa Lohani

13 Pharmacokinetic and Molecular Docking Studies of Plant-Derived Natural Compounds to Exploring Potential Anti-Alzheimer Activity 217
Aftab Alam, Naaila Tamkeen, Nikhat Imam, Anam Farooqui, Mohd Murshad Ahmed, Safia Tazyeen, Shahnawaz Ali, Md Zubbair Malik, and Romana Ishrat

14 In Silico Biochemical Pathways for Bacterial Metabolite Synthesis 239
Murtaza Ali, Shahnawaz Ali, and Romana Ishrat

15 In Silico Characterization of Plant Secondary Metabolites 251
A. Sabitha Rani, G. Neelima, Rupasree Mukhopadhyay, K. S. N. Jyothi, and G. Sulakshana

16 In Silico Identification of Plant-Derived Secondary Metabolites in Defense 275
Kalaivani K. Nadarajah

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

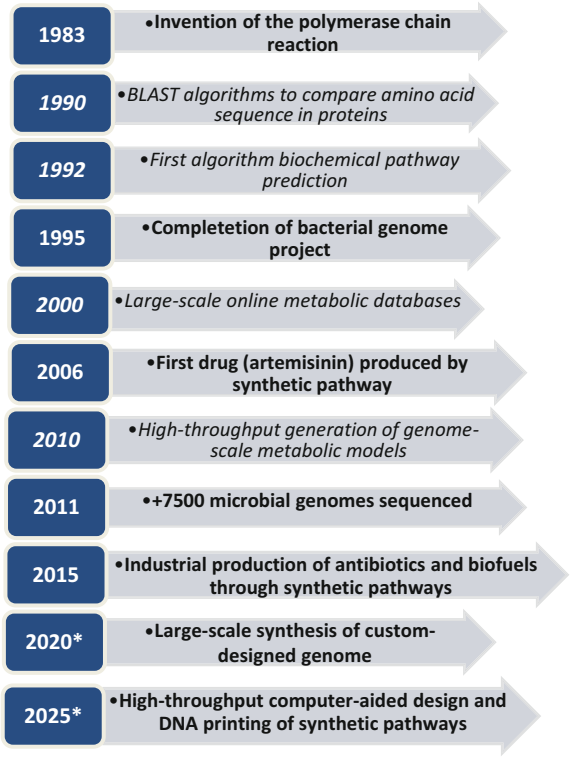


Fig. 1.1 Time scale of the main events for the evolution of de novo synthesis and pathway engineering. The evolution based on computational analysis is highlighted in italics, and the wet-lab-based evolution is shown in *bold letters*. *Future developments

"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,