Contents

List of Contributors *xxi*
Preface *xxvii*

1 **Bioinformatics as a Powerful Tool to Foster Plant Science Research and Crop Breeding Through Its Involvement in a Multidisciplinary Research Activity** *1*
Jemaa Essemine, Zhan Xu, Jen-Tsung Chen, and Mingnan Qu
1.1 Introduction *1*
1.2 Bioinformatics as a Powerful Tool for Big Data Analysis in Plant Science *3*
1.3 Role of Bioinformatics in Trait Mapping *3*
1.4 Bioinformatics in Molecular Biology *3*
1.5 Role of Bioinformatics in Genetic Variation *4*
1.6 Bioinformatics in Genome-wide Association Studies (GWAS) *4*
1.7 Implication of Bioinformatics in "Omics" *5*
1.8 Bioinformatics in Computational Biology and Evolutionary Studies *5*
1.9 Role of Bioinformatics in Transcriptomics *6*
1.10 Implication of Bioinformatics in Next-generation Sequencing (NGS) Analysis *6*
1.11 Implication of Bioinformatics in Metabolomics *7*
1.12 Bioinformatics and Epigenetics *8*
1.13 Involvement of Bioinformatics in Synthetic Biology *9*
1.14 How Can Bioinformatics Promote Plant Biotechnology? *9*
1.15 Bioinformatics Use in Biotic and Abiotic Stress Management *10*
1.16 Bioinformatics for the Investigation of Plant Resistance to Pathogens *11*
1.17 Bioinformatics in Crop Breeding and Improvement *12*
1.18 Bioinformatics Impacts on Plant Science *13*
1.19 Application of Bioinformatics in Plant Breeding Programs *13*
1.20 Conclusion *14*
References *15*

2 **Bioinformatics for Molecular Breeding and Enhanced Crop Performance: Applications and Perspectives** *21*
Rahul Lahu Chavhan, Vidya Ramesh Hinge, Dipti Jayvantrao Wankhade, Abhijeet Subhash Deshmukh, Nagrani Mahajan, and Ulhas Sopanrao Kadam
2.1 Introduction *21*
2.1.1 Bioinformatics and Its Application in Plant Breeding *21*

2.1.2 Applications of Bioinformatics in Plant Breeding *22*
2.2 Data Management and Integration *22*
2.2.1 Data Storage and Retrieval *22*
2.2.2 Data Integration and Analysis *24*
2.2.3 Data Visualization and Interpretation *25*
2.3 Genomic Resources for Plant Breeding *26*
2.3.1 Plant Genome Sequencing and Annotation *26*
2.3.2 Transcriptome Analysis and Gene Expression Profiling *27*
2.3.3 Proteome Analysis and Protein–Protein Interactions *42*
2.3.4 Metabolome Analysis and Metabolic Pathway Reconstruction *43*
2.3.4.1 Metabolic Pathway Reconstruction *44*
2.3.4.2 Metabolic Pathway Databases *45*
2.4 Application of Bioinformatics, Genomics, and Proteomics in Crop Improvement and Breeding *45*
2.4.1 Bioinformatics Databases and Tools in Genomic Research *45*
2.4.2 Genetic Diversity Study and Estimation of Phylogenetic Relationships *46*
2.4.3 Genomic Selections and their Application in the Breeding Program *46*
2.4.3.1 Role of Genomic Selections (GS) in Identifying Genes and Proteins Associated with Desirable Traits *51*
2.4.4 SNP Discovery and SNP Genotyping *51*
2.4.5 Breeding with Marker Assistance to Improve Crops *52*
2.4.6 Association Mapping Studies in Plants *53*
2.4.7 Mapping of QTL *54*
2.4.8 Linkage Mapping of Molecular Markers and Oligogenic Traits *55*
2.4.9 Role of Genomics Tools such as MAS, GWAS, and Genomic Selections in Identifying Genes and Proteins Associated with Desirable Traits *55*
2.4.9.1 Genome-wide Association Studies in Identifying Proteins and Genes Linked to Desired Traits *57*
2.4.10 Study of Entire Genomes and Identification of Desirable Genetic Traits *59*
2.4.11 Genomics, Transcriptomics, Proteomics, and Metabolomics Studies in Marker and Trait Identification *60*
2.4.12 Genome Editing Application in Crop Improvement *60*
2.5 Challenges and Future Directions *61*
2.5.1 Interpretation and Management of Large Genetic Data as a Challenge *61*
2.5.2 Need for Advanced Methods and Novel Bioinformatics Tools *62*
2.5.3 Potential for Bioinformatics to Contribute to Sustainable Agricultural Development *62*
2.6 Conclusions *63*
References *63*

3 Multi-omics: An Advanced Bioinformatics Approach for Crop Improvement in Agriculture *75*
Vinay Kumar Dhiman, Devendra Singh, Vivek Kumar Dhiman, and Himanshu Pandey
3.1 Multi-omics: A Boon to Crop Improvement *75*
3.2 Genomics: Unlocking the Crop Genome *77*

3.2.1 Plant Genome Databases *79*
3.2.2 Functional Genomics *82*
3.2.2.1 Functional Genomics Using Forward and Reverse Genetics *82*
3.2.2.2 Functional Genomics in Relation to DNA *84*
3.2.2.3 Functional Genomics in Relation to RNA and Protein *85*
3.2.3 Structural Genomics *87*
3.2.4 Comparative Genomics *88*
3.2.5 Epigenomics *88*
3.3 Metabolomics: Profiling the Crop's Metabolic Processes *89*
3.4 Phenomics *90*
3.5 Ionomics *92*
3.6 Omics-Assisted Breeding: Accelerating Crop Improvement *92*
3.7 Conclusion and Future Perspectives *92*
References *94*

4 Genetic Mapping of Valued Genes with Significant Traits in Crop Plants: Basic Principles, Current Practices, and Future Perspectives *99*
Prasanta Kumar Majhi, Akansha Guru, Suma C. Mogali, Prachi Pattnaik, Ritik Digamber Bisane, Lopamudra Singha, Partha Pratim Behera, and Prateek Ranjan Behera
4.1 Introduction *99*
4.2 Quantitative Trait Loci (QTLs) and Genetic Mapping of Traits *101*
4.2.1 Objectives of QTL Mapping *102*
4.3 The Fundamentals of the QTL Mapping Approach *102*
4.3.1 Salient Requirements of QTL Mapping *103*
4.3.2 Basic Method for QTL Analysis *103*
4.3.3 Conventional QTL Mapping (Linkage Mapping) *104*
4.4 Mapping Populations Used in QTL Mapping Experiments *104*
4.4.1 Parent Selection for Mapping Population Development *105*
4.4.2 Family-based Mapping Populations *105*
4.4.3 Natural Population-based Mapping *113*
4.4.4 Bulked Segregant Analysis (BSA) *116*
4.4.5 Choice of Appropriate Mapping Population: Types and Size *117*
4.5 Molecular Markers for QTL Mapping *119*
4.6 Statistical Approaches for Detection of QTLs *121*
4.6.1 Single-marker Analysis (Point Analysis) *121*
4.6.2 Interval Mapping by Maximum Likelihood *122*
4.6.3 Interval Mapping by Regression *122*
4.6.4 Simple Interval Mapping (SIM) *123*
4.6.5 Composite Interval Mapping (CIM) *123*
4.6.6 Multiple Interval Mapping (MIM) *124*
4.7 Software Used for QTL Mapping *124*
4.8 QTLs and the Signature of Selection *125*
4.9 Factors Affecting the Power of QTL Mapping *125*
4.10 Merits of QTL Mapping *128*

4.11 Demerits of QTL Mapping *128*
4.12 Conclusion and Way Forward *129*
References *130*

5 **Basic Bioinformatics for Identification and Analysis of Candidate Genes in Plants Toward Crop Improvement** *135*
Sadhana Singh
5.1 Introduction *135*
5.2 Candidate Genes such as Transcription Factors and Gene Families *137*
5.2.1 Importance of NAC Transcription Factors *137*
5.2.2 Importance of DREBs/CBFs *138*
5.2.3 Availability of Draft Genome Sequences *139*
5.3 Methods *140*
5.3.1 Identifying and Examining Genes in Crop Species to Understand their Characteristics through Data Analysis *143*
5.3.2 Transmembrane Domains Prediction and Orthologs Distribution *146*
5.3.3 Phylogenetic Analysis and Identification of Putative Stress-responsive NAC Genes *147*
5.3.4 Conserved Motifs, and Gene Structure Analysis of Putative Stress-responsive Genes (NAC) *147*
5.3.5 Cis-acting Regulatory Elements in Stress-responsive Genes Identification *150*
5.3.6 STRING Evaluation for Protein–Protein Network Analysis *151*
5.3.7 *In silico* Expression Analysis of NAC Genes in Food Legumes *152*
5.3.8 Quantitative PCR (Real Time) Analysis *154*
5.4 Conclusion *154*
References *155*

6 **Exploring Machine Learning Algorithms for Gene Function Prediction in Crops** *159*
Ruchi Jakhmola-Mani, Sonali, Aniket Pandey, Dhananjay Raturi, Rishita Singh, Kusala Vanam, Manish D, Ritu Chauhan, Deepshikha Pande Katare, Potshangbam Nongdam, and Angamba Meetei Potshangbam
6.1 Introduction *159*
6.1.1 Why is the Function of Genes Important? *159*
6.1.2 Genetics and Crop Improvement in the Past *160*
6.1.3 The Impact of Precise Gene Location/Function Prediction on Crop Improvement *161*
6.1.4 Importance of Crop Improvement in the 21st Century *161*
6.1.5 Global Food Demand, Changing Dietary Habits Challenges *163*
6.1.6 Need for Efficient and Faster Crop Improvement Methods *163*
6.1.7 Multidisciplinary Approach Combining Genetics, Plant Breeding, Biotechnology, and Data Science for New Crop Improvement Strategies *164*
6.2 Computational Methods for Gene Function Prediction *164*
6.2.1 Bioinformatics *164*
6.2.1.1 Bioinformatics Tools in Plant Disease Management *165*

6.2.2 Machine Learning *166*
6.2.2.1 Supervised Learning *166*
6.2.2.2 Unsupervised Learning *166*
6.3 Machine Learning and Crop Improvement *167*
6.3.1 Dataset *168*
6.3.1.1 Data Integration *168*
6.3.1.2 Splitting the Dataset *168*
6.3.2 Overview of Machine Learning Algorithms *169*
6.3.2.1 k-Nearest Neighbors *169*
6.3.2.2 Random Forest *169*
6.3.2.3 Principal Component Analysis *171*
6.3.2.4 Deep Learning Algorithms *171*
6.3.3 Performance Assessment *172*
6.4 Experiment *173*
6.4.1 Dataset Description *174*
6.4.1.1 Training and Test Set *174*
6.4.1.2 File Description *174*
6.4.1.3 Protein Embeddings for Train and Test Set *174*
6.4.2 Training *175*
6.4.2.1 Model Description *175*
6.4.2.2 Hyperparameters *175*
6.4.3 Result *175*
6.5 Case Studies and Success Stories *176*
6.5.1 Highlighting Successful Applications of Machine Learning in Crop Improvement *176*
6.5.2 Examples of Improved Crop Traits Through Gene Function Prediction *177*
6.6 Challenges and Future Directions *178*
6.6.1 Addressing Challenges and Limitations of Machine Learning in Crop Improvement *178*
6.6.1.1 Limited Availability of Data *178*
6.6.1.2 Learning Algorithm *179*
6.6.1.3 Multidisciplinary Expertise Gap *179*
6.6.2 Prospects and Potential Advancements in the Field *179*
6.6.2.1 ML for Gene Function Prediction in Large Multi-crop Datasets *179*
6.6.2.2 Revealing Potential Target for Crop Improvement *180*
6.6.2.3 Novel Gene Function Prediction *180*
References *180*

7 Omics and Bioinformatics Approaches for Abiotic Stress Tolerance in Plants *185*
Santanu Samanta and Aryadeep Roychoudhury
7.1 Introduction *185*
7.2 Genomic Approaches *186*
7.2.1 Structural Genomics *186*
7.2.1.1 Genome Sequencing *186*

7.2.1.2 Molecular Mapping *187*
7.2.1.3 Molecular Markers *187*
7.2.2 Functional Genomics *188*
7.2.2.1 Expressed Sequence Tags (ESTs) *188*
7.2.2.2 Serial Analysis of Gene Expression (SAGE) *189*
7.2.2.3 Massively Parallel Signature Sequencing (MPSS) *189*
7.3 Transcriptomics Approaches *189*
7.3.1 Microarray and Gene Chip *190*
7.3.2 RNASequencing (RNA-seq) *190*
7.4 Proteomics Approaches *191*
7.4.1 Gel-based Techniques *191*
7.4.1.1 Two-dimensional Gel Electrophoresis (2-DGE) *191*
7.4.1.2 Two-dimensional Difference in Gel Electrophoresis (2-DDIGE) *192*
7.4.2 Gel-free Techniques *192*
7.4.2.1 Multidimensional Protein Identification Technology (MudPIT) *192*
7.4.2.2 Isotope-coded Affinity Tags (ICAT) *193*
7.4.2.3 Isobaric Tags for Relative and Absolute Quantification (iTRAQ) *193*
7.4.2.4 Stable Isotopic Labeling with Amino Acids in Cell Culture (SILAC) *194*
7.5 Metabolomics Approaches *194*
7.5.1 Gas Chromatography/MS (GC-MS) *195*
7.5.2 Liquid Chromatography-MS (LC-MS) *195*
7.5.3 Capillary Electrophoresis-MS (CE-MS) *195*
7.5.4 Fourier Transform Ion Cyclotron Resonance-MS (FT-ICR-MS) *196*
7.6 Bioinformatics Approaches *196*
7.6.1 Biological Databases *196*
7.6.2 Bioinformatics Tools and Software *197*
7.7 Concluding Remarks *197*
Acknowledgments *198*
References *198*

8 **Bioinformatics Approaches for Unraveling the Complexities of Plant Stress Physiology** *209*
Sneha Murmu, Himanshushekhar Chaurasia, Ipsita Samal, Tanmaya Kumar Bhoi, and Asit Kumar Pradhan
8.1 Introduction *209*
8.2 Understanding Plant Stress Response Mechanisms *210*
8.2.1 Overview of Plant Stress Signaling Pathways *210*
8.2.2 Role of Stress-responsive Genes and Proteins *210*
8.2.3 Regulatory Networks Involved in Plant Stress Response *211*
8.3 Genome and Transcriptome Analysis for Plant Stress Physiology *212*
8.3.1 Genomics *212*
8.3.2 Transcriptomics *214*
8.4 Proteomics and Metabolomics Approaches *216*
8.4.1 Proteomics *216*
8.4.2 Metabolomics *218*
8.5 Data Integration and Systems Biology Approaches *220*

8.6 Bioinformatics Resources for Plant Stress *221*
8.6.1 Databases *221*
8.6.2 Machine Learning Tools *222*
8.7 Conclusion *226*
References *226*

9 Bioinformatics Tools for Assessing Drought Stress Tolerance in Crops *233*
Nageswara Rao Reddy Neelapu and Kolluru Viswanatha Chaitanya
9.1 Introduction *233*
9.2 Bioinformatics for Plant Research and Crop Breeding *234*
9.3 Genomics and Drought Stress Tolerance *234*
9.3.1 Genome-wide Association Studies (GWAS) *235*
9.4 Transcriptome Analysis for the Drought Stress Tolerance *236*
9.4.1 MicroRNAs (miRNAs) and Drought Stress *238*
9.5 Proteome and Drought Stress *239*
9.6 Metabolomics and Drought Stress Tolerance *241*
9.7 Phenome and Drought Stress *242*
9.8 Future of the Omics technologies *244*
9.9 Conclusions *245*
References *246*

10 Bioinformatics Tools and Resources for Plant Transcriptomics: Challenges and Opportunities *251*
Sona Charles and Meriin Lopus
10.1 Introduction *251*
10.2 Evolution of Transcriptomic Technologies *252*
10.3 Steps in Transcriptomic Data Analysis *254*
10.3.1 Raw Data Preprocessing *254*
10.3.2 Genome Assembly and Mapping *254*
10.3.3 Differential Expression Analysis *255*
10.3.4 Functional Enrichment Analysis *257*
10.3.5 Pathway Analysis *257*
10.3.6 GO Analysis *259*
10.4 R/Bioconductor Packages for Transcriptomic Analysis *259*
10.5 Galaxy Server for Transcriptome Analysis *260*
10.6 Stress Transcriptomics – A Case Study *260*
10.7 Conclusion and Way Forward *262*
References *262*

11 Development of a Core Set from Large Germplasm Collections in Genebank *269*
Pradeep Ruperao
11.1 Introduction *269*
11.2 Developing a Core Collection *270*
11.3 Constructing a Core Collection *270*

11.3.1 Single Nucleotide Polymorphisms (SNPs) *270*
11.3.2 k-mer Analysis *271*
11.3.3 Heterozygosity *272*
11.3.4 Genetic Distance *273*
11.3.5 Available Tools *273*
11.4 Assessing the Core Collections *275*
11.4.1 Background Information *275*
11.4.2 Diversity Analysis *275*
11.4.3 Phenotypic Characterization *276*
11.4.4 Genetic Marker Analysis *276*
11.4.5 Phylogenetic Analysis *276*
11.4.6 Trait Performance Assessment *276*
11.4.7 Evaluate Representativeness *276*
11.4.8 Allele Coverage *276*
11.4.9 Heterozygosity *277*
11.4.10 Shannon Diversity *277*
11.4.11 Genetic Distance *277*
11.4.12 Agro-climatic Conditions *278*
11.5 Conclusion and Future Considerations *278*
11.5.1 Dynamic Updating *278*
11.5.2 Integration of Multi-omics Data *279*
11.5.3 Climate Resilience *279*
11.5.4 Genomic Selection and Breeding Applications *279*
11.5.5 Precision Conservation *279*
11.5.6 Community Engagement and Indigenous Knowledge *279*
11.5.7 Digital Repositories and Data Sharing *279*
11.5.8 Crop Wild Relatives *279*
11.5.9 Machine Learning and AI *279*
11.5.10 International Collaboration *280*
11.5.11 Education and Awareness *280*
11.5.12 Food Security and Nutrition *280*
11.5.13 Crop Disease and Pest Resistance *280*
References *280*

12 **Bioinformatics Approaches to Determine Plant microRNA Targets** *283*
Shree Prakash Pandey
12.1 Introduction *283*
12.2 Characteristic Features and Principles of miRNA-targeting in Plants *285*
12.2.1 Sequence Complementarity *285*
12.2.2 Target Accessibility and Flanking Sequences *286*
12.2.3 Conservation *287*
12.2.4 Expression Patterns *287*
12.3 Tools for miRNA Target Prediction in Plants *288*
12.4 Bioinformatics Identification of miRNA and mRNA at a Genome-scale *291*

12.5 Conclusion 292
References 293

13 **Machine Learning for the Discovery of DNA-binding Proteins in Plants** 299
Upendra Kumar Pradhan, Prabina Kumar Meher, and Pushpendra Kumar Gupta
13.1 Introduction 299
13.2 Steps Involved in Identification of DBPs Using Machine Learning 301
13.2.1 Collection and Processing of Sequence Data 301
13.2.2 Extraction of Numeric Features from Sequence Data 302
13.2.2.1 Sequence-derived Features 303
13.2.2.2 Structure-based Features 305
13.2.2.3 Physico-chemical Features 305
13.2.3 Prediction Using Machine Learning Algorithms 305
13.2.3.1 Shallow Learning Models 306
13.2.3.2 Deep Learning Models 309
13.3 Assessment of Learning Algorithms for DBP Prediction Using Sequence- and PSSM-derived Features 311
13.4 Evaluation of Existing Tools for DBP Prediction in Plants 313
13.5 Conclusion and Future Perspectives 314
References 315

14 **Bioinformatics for Gene Identification and Crop Improvement in Wheat** 321
Pushpendra Kumar Gupta, Jyoti Chaudhary, and Tinku Gautam
14.1 Introduction 321
14.2 Databases and Tools for Individual Genes and Proteins 321
14.3 Identification/Characterization of Genes/Gene Families at the DNA Level 325
14.3.1 Identification of Genes 325
14.3.1.1 Identification of Genes by Similarity-based Search and Sequence Retrieval 325
14.3.1.2 Identification of Genes through Ab Initio Prediction 326
14.3.1.3 Identification of Alien Genes in Wheat Cultivars 327
14.3.2 Characterization of Retrieved Gene Sequences 327
14.4 Characterization of Genes at the Protein Level 328
14.4.1 Physico-chemical Properties and Subcellular Location 328
14.4.1.1 Physico-chemical Properties 328
14.4.1.2 Subcellular Localization 329
14.4.2 3D Structure 329
14.4.3 Protein–Protein Interactions (PPI) 330
14.4.4 Molecular Docking and Molecular Dynamics (MD) Simulations 330
14.5 Phylogenetic Analysis 331
14.6 Present Status of Wheat Genes Identified *in silico* 331
14.7 Utility of Predicted Genes for Crop Improvement 337
14.7.1 Molecular Breeding Involving Molecular Markers 337
14.7.2 Gene Editing 339
14.7.3 Transgenic Wheat 339

14.8 Conclusion and Prospects *340*
References *340*

15 Bioinformatics for Analyzing the Role of Epigenetics in Plant Disease Resistance *351*
Kalpana Singh, Harindra Singh Balyan, and Pushpendra Kumar Gupta
15.1 Introduction *351*
15.2 Histone Modifications *351*
15.2.1 Types of Histone Modifications (Acetylation, Methylation, Ubiquitylation) *351*
15.2.1.1 Histone Acetylation *352*
15.2.1.2 Histone Methylation *353*
15.2.1.3 Histone Ubiquitination *353*
15.2.2 Generation of ChIP-seq/CUT&Tag Data for the Study of Histone Modification *354*
15.2.3 Bioinformatic Analysis of ChIP-seq and CUT&Tag Data *355*
15.2.3.1 Checking Quality of DNA Sequences *355*
15.2.3.2 Elimination of Low-quality Reads and Bases *356*
15.2.3.3 Correlation Studies *356*
15.2.3.4 Retrieval of Histone Modification Sites *356*
15.2.3.5 Reproducibility Analysis *356*
15.2.3.6 Retrieval of Histone Binding Sites *357*
15.2.3.7 Genomic and Functional Annotation of Histone binding Sites *357*
15.3 Chromatin Accessibility *357*
15.4 DNA Methylation *360*
15.4.1 DNA Methylation Sites in the Genome *360*
15.4.2 Methylation/Demethylation Enzymes *360*
15.4.3 Role of DNA Methylation in Disease Resistance *360*
15.4.4 Techniques and Data Analysis for the Study of DNA Methylation *361*
15.4.4.1 MSAP and MSAP-seq *361*
15.4.4.2 MeDIP-seq *362*
15.4.4.3 BS-seq (WGBS/BS-seq and RRBS-seq) *362*
15.4.4.4 Software for Methylated Transposable Elements (TEs) *364*
15.5 Noncoding RNAs (miRNAs, lncRNA, circRNA) *365*
15.5.1 Small RNA Sequencing and miRNAs *366*
15.5.2 Long ncRNA (lncRNAs) *368*
15.5.3 Circular RNAs *369*
15.6 Conclusions and Future Perspectives *370*
References *371*
Weblinks *390*

16 The Evolution of Auxin-Binding Protein 1 *391*
Siarhei A. Dabravolski and Stanislav V. Isayenkov
16.1 Abundance of Auxin and Auxin-binding Proteins in Nature *391*
16.2 Auxin in Plants *392*
16.3 Domain Organization *393*

16.4 ABP1 Active Sites/Structure/Sequence Analysis *395*
16.4.1 Homodimerisation and Glycosylation Sites *397*
16.4.2 KDEL *397*
16.4.3 Zn-binding Sites *397*
16.4.4 Auxin-binding Sites *397*
16.5 ABP1 Evolution *398*
16.5.1 Phylogenetic Analysis *398*
16.5.2 Structures Modelling and Comparison *400*
16.6 Future Prospective *403*
16.6.1 Role in Archaea *403*
16.6.2 Homologous to Bacterial Lys-type Transcriptional Regulator (LTTRs) Auxin-binding Proteins in Plants *403*
16.6.3 Function of Other Cupin Domain-containing Proteins in Plants? *403*
16.6.4 How Hormone-binding Efficiency/Selectivity has changed Over Evolution? *404*
16.6.5 IAA Sensors in Fungi, Insects, and Animals? *404*
16.6.6 Toxic Effect of IAA on Animals *404*
16.6.7 Positive Effects of IPA/IAA on Animals *405*
16.7 Conclusion *405*
References *405*

17 Exploring the Potential of Molecular Docking and *In Silico* Studies in Secondary Metabolite and Bioactive Compound Discovery for Plant Research *413*
Amine Elbouzidi, Mohamed Taibi, and Mohamed Addi
17.1 Introduction *413*
17.2 Importance of Structure-based Drug Design from Natural Sources *415*
17.3 Molecular Docking as a Key Component of SBDD: A Bridge Between Computational and Experimental Approaches 417
17.4 Molecular Docking and Natural Product Database *419*
17.4.1 Constructing and Preparing *In Silico* Ligands *421*
17.4.2 Preparing the Target *421*
17.4.3 Docking *421*
17.4.4 Identifying Hit Molecules *421*
17.4.5 Optimization of Hits *423*
17.5 Case Studies: Successful Applications of *In Silico* Molecular Docking in Plant Research for Diverse Applications *423*
17.6 Concluding Remarks and Future Considerations *429*
References *430*

18 Exploring Secondary Metabolites in Plants Through Bioinformatics *435*
Sneha Murmu, Ritwika Das, Bharati Pandey, Soumya Sharma, and Mohammad Samir Farooqi
18.1 Introduction *435*
18.2 Classification of Plant Secondary Metabolites *436*
18.2.1 Terpenes *438*
18.2.2 Phenolics *438*

18.2.3 N-containing Compounds *438*
18.2.4 S-containing Compounds *438*
18.3 Secondary Metabolites Pathways in Plants *438*
18.3.1 Shikimic Acid Pathway *439*
18.3.2 Terpene Biosynthesis Pathway *439*
18.3.3 Phenylpropanoid Pathway *439*
18.3.4 Alkaloid Biosynthesis Pathway *440*
18.3.5 Glucosinolate Biosynthesis Pathway *440*
18.3.6 Glycoside Biosynthesis Pathway *440*
18.4 Mining of Omics Data *440*
18.4.1 Genomics *441*
18.4.1.1 Genomic Data Retrieval *441*
18.4.1.2 Genome Annotation *441*
18.4.1.3 Comparative Genomics *442*
18.4.1.4 Differential Expression Analysis *442*
18.4.1.5 Functional Annotation *443*
18.4.1.6 Co-expression Analysis *443*
18.4.2 Transcriptomic Data *443*
18.4.2.1 Data Acquisition *443*
18.4.2.2 Data Preprocessing *443*
18.4.2.3 RNA-Seq Data Alignment *444*
18.4.2.4 Annotation of Transcriptomic Data *444*
18.4.3 Proteomics Data *444*
18.4.3.1 Data Collection *445*
18.4.3.2 Data Repositories and Format Conversion *445*
18.4.3.3 Quality Control *445*
18.4.3.4 Data Preprocessing *445*
18.4.3.5 Protein Identification and Quantification *445*
18.4.3.6 Differential Expression Analysis *445*
18.4.4 Metabolomics Data *446*
18.4.4.1 Data Acquisition *446*
18.4.4.2 Feature Identification *446*
18.4.4.3 Differential Analysis *446*
18.4.4.4 Functional Annotation *446*
18.4.5 Multi-omics Integration *446*
18.5 Bioinformatics Tools for Analysis of Secondary Metabolites and Pathways *447*
18.5.1 Data Processing and Annotation *447*
18.5.2 Pathway Analysis *450*
18.6 Conclusion *452*
References *452*

19 Understanding Plant Secondary Metabolism Using Bioinformatics Tools: Recent Advances and Prospects *459*
Dola Mukherjee and Ashutosh Mukherjee
19.1 Introduction *459*

19.2 Secondary Metabolic Gene Clusters *461*
19.3 Sequencing Techniques and Analytical Tools for Plant Metabolomics Study *462*
19.4 Bioinformatics Tools for the Elucidation of Secondary Metabolism in Plants *465*
19.5 Medicinal Plant Genome and/or Metabolome Databases *465*
19.5.1 1 K Medicinal Plant Genome Database *465*
19.5.2 TCMPG *468*
19.5.3 IGTCM *468*
19.5.4 IMPPAT *468*
19.5.5 GPGD *469*
19.5.6 CMAUP *469*
19.6 Automation of Natural Product Detection by Identification of Metabolic Gene Cluster *469*
19.7 The Big Data and Systems Biology Approach *470*
19.8 Application of Machine Learning in Plant Secondary Metabolism *471*
19.9 Artificial Intelligence (AI) *472*
19.10 Machine Learning (ML) *472*
19.11 Deep Learning (DL) *473*
19.11.1 Convolutional Neural Network (CNN) *474*
19.11.2 Recurrent Neural Network (RNN) *474*
19.11.3 Long Short-Term Memory (LSTM) *474*
19.11.4 Generative Adversarial Network (GAN) *474*
19.11.5 Artificial Neural Networks (ANNs) *475*
19.12 Conclusion and Future Perspective *475*
References *477*

20 **An Appraisal of Flavonoids Through Bioinformatics** *489*
Manoj Kumar Mishra and Vibha Pandey
20.1 Overview of Flavonoids *489*
20.2 Identification of Flavonoid Biosynthetic Genes and Enzymes by Computational Tools *490*
20.2.1 Chemoinformatics and Molecular Docking *491*
20.2.2 Virtual Screening *491*
20.2.2.1 Virtual High-throughput Screening *491*
20.2.2.2 Active Site Prediction *492*
20.2.2.3 Structure-based Screening *492*
20.2.2.4 Ligand-based Screening *492*
20.2.2.5 Machine Learning Techniques *492*
20.2.2.6 Virtual Combinatorial Libraries *493*
20.2.3 Flavon Q *493*
20.2.4 Flavon Q-2.0 *493*
20.2.5 Bioinformatics Tools *493*
20.3 Prediction of the Potential Biological Activities of Flavonoids Based on Their Chemical Structure *494*
20.3.1 Quantitative Structure–Activity Relationship (QSAR) *494*

20.3.2 Machine Learning and Data Mining 495
20.3.3 Molecular Dynamics Simulation 495
20.4 Chalcone Synthase 495
20.5 Sequence Retrieval 496
20.5.1 Through Cloning 496
20.5.2 Online Search 496
20.6 Localization 497
20.7 Homology Search 497
20.8 Conserved Domain 497
20.9 Sequence Alignment and Phylogeny 498
20.9.1 Active Site 498
20.9.2 Phylogeny 498
20.10 Chromosome Location 499
20.11 Characterization 500
20.12 Three-dimensional Structure 500
20.13 Docking 501
20.14 Conclusion and Prospects 501
References 501

21 **Golden Opportunities: Harnessing Bioinformatics to Revolutionize Plant Research and Unleash the Power of Golden Rice in Crop Breeding** 505
Poulami Majumder
21.1 Introduction 505
21.1.1 An Overview of Bioinformatics in Plant Research and Crop Breeding 507
21.1.2 The Significance of Golden Rice in Global Agriculture 508
21.1.3 Objective and Scope of the Chapter 509
21.2 Background 510
21.3 Bioinformatics Tools and Resources for Plant Research 512
21.3.1 Sequence Databases and Repositories 512
21.3.2 Sequence Alignment and Phylogenetic Analysis 513
21.3.2.1 Applications in Plant Genomics 514
21.3.3 Functional Genomics and Transcriptomics Tools 514
21.3.4 Applications in Golden Rice 516
21.3.4.1 Integration of Omics Data for Systems Biology 516
21.4 Genetic Engineering and Breeding Strategies for Golden Rice 517
21.4.1 CRISPR-Cas9 and Other Gene-editing Technologies 519
21.4.2 Marker-assisted Selection (MAS) 520
21.4.2.1 Advantages of MAS 520
21.4.2.2 Limitations 520
21.4.2.3 Applications 520
21.4.3 Genomic Selection and Genome-wide Association Studies (GWAS) 521
21.4.4 Speed Breeding and High-throughput Phenotyping 522
21.5 Case Study: Development and Improvement of Golden Rice 523
21.5.1 The Initial Development of Golden Rice (GR1 and GR2) 523
21.5.1.1 Challenges and Regulatory Hurdles 524

21.5.1.2 Challenges and Continued Development *524*
21.6 Bioinformatics-guided Identification of Target Genes for Provitamin A Enhancement *525*
21.6.1.1 Employing CRISPR-Cas9 and Other Breeding Strategies for Golden Rice Improvement *526*
21.6.2 Assessing the Agronomic Performance and Nutritional Quality of Improved Golden Rice Lines *527*
21.7 Ethical and IP Issues *529*
21.7.1 Ethical Considerations and Public Perception of Genetically Modified Crops *529*
21.7.2 Regulatory and Intellectual Property Issues *531*
21.7.3 Ensuring Accessibility and Adoption of Golden Rice in Target Communities *532*
21.8 Conclusion *533*
Declaration *534*
References *535*

22 **Going Wild: Genomics of Forest Plants and the Future of Crop Improvement** *539*
Ajinkya Bharatraj Patil, Debojyoti Kar, Sourav Datta, and Nagarjun Vijay
22.1 Introduction *539*
22.2 Repetitive Genomic Elements as Phenotypic Trait Variation Machinery *542*
22.3 Naturally Acquired Traits Can Be Effectively Characterized Using Population Genomics *545*
22.4 Demographic Forces Are Crucial in Shaping Genome Dynamics *549*
22.5 Conclusion *552*
References *552*

Index *559*