

Contents

Chapter 1

Rice Genome Sequencing and Data Mining Resources 1

Baltazar A. Antonio, Yoshiaki Nagamura, Nobukasu Namiki,
Takashi Matsumoto, and Takuji Sasaki

1. Introduction	1
2. Rice Genome Sequencing	2
3. Rice Genome Informatics	4
4. Functional Genomics Resources	6
5. Towards an Integrated Data Mining Resource	8
6. References	9

Chapter 2

Application of Evolutionary Computation to Bioinformatics 13

Daniel Ashlock

1. Introduction	13
2. Evolutionary Computation	14
3. Finding a Transposon Insertion Motif	16
4. PCR Primer Picking	18
5. DNA Bar Code Location	24
6. Summary	28
7. References	29

*Chapter 3***Architectures for Integration of Data and Applications: Lessons from Integration Projects** 31

William D. Beavis

1. Introduction	31
2. A Web-Accessible Data Warehouse (TAIR)	35
2.1. Description	35
2.2. Lessons	37
3. An Automated Analysis and Annotation Pipeline (XGI).....	37
3.1. Description	39
3.2. Lessons	40
4. An Open Source Integrated System (Genex-Lite).....	40
4.1. Introduction	43
4.2. Lessons	43
5. Loose Coupling Integration (ISYS)	43
5.1. Description	45
5.2. Lessons	45
6. Summary of Lessons Learned from All Integration Projects	45
7. References	46

*Chapter 4***Functional Genomics Approach to Elucidate the Regulation of Vascular Development in Poplar** 49

Rishikesh P. Bhalerao and Göran Sandberg

1. Introduction	49
2. Poplar as a Model for Studying Formation	51
3. Elucidating the Control of Secondary Xylem Development.....	51
3.1. The Role of Aux/IAA Genes in Regulation of Secondary Xylem Development.....	51
3.2. Genomics Approach to Elucidation of Secondary Xylem Development	53
4. The Seasonal Control of Cambial Activity	56
4.1. Large Scale Sequencing of Active and Dormant Cambial Libraries	57
4.2. Transcript Profiling of Active and Dormant Cambium.....	58
5. Future Perspectives.....	59
5.1. Arabidopsis as Test Bed for Rapidly Identifying Genes Involved in Vascular Development	59
5.2. Rapid Analysis of Gene Function of Poplar Genes with Unknown Function	60

6. Conclusions	60
7. References	61

Chapter 5

Novel Tools for Plant Genome Annotation and Applications to <i>Arabidopsis</i> and Rice	63
--	----

Volker Brendel

1. Introduction	63
2. Boon and Bane of the Industrialization of Molecular Biology and Genetics	64
3. The Gene Structure Annotation Problem	65
4. Gene Structure Annotation by Spliced Alignment	65
5. The Case for a Community Approach to Annotation	68
6. Perspective	75
7. References	75

Chapter 6

FCModeler: Dynamic Graph Display and Fuzzy Modeling of Metabolic Maps	77
--	----

Julie A. Dickerson

1. Modeling Metabolic Networks	77
2. Metabolic Networking Database	78
3. Modeling Metabolic Relationships	79
3.1. Graphing the Metabolic Map: FCModeler	80
3.2. Modeling Metabolic Networks Using Fuzzy Cognitive Maps	81
3.3. Cycle Analysis in Metabolic Networks	82
4. Representing Metabolic Pathways in Virtual Reality	82
4.1. 3-D Network Visualization	83
4.2. User Interaction and Navigation	83
4.3. Virtual Cell Representation	85
5. Summary	85
6. References	86

Chapter 7

Old Methods for New Ideas: Genetic Dissection of the Determinants of Gene Expression Levels	89
--	----

Kyunga Kim, Marilyn A.L. West, Richard W. Michelmore,
Dina A. St. Clair, and R.W. Doerge

1. Introduction	89
2. Genetical Genomics	90
3. Methods	91
3.1. Dye Swap Experimental Design	92
3.2. Modelling Determinants of Expression Level Polymorphisms	92
3.3. Experimental Design for Expression Level Polymorphism Analysis	94
3.4. Implementation of ELP Mapping Using Existing Multiple Trait Mapping Methodology	96
4. Simulation Study	97
4.1. Simulation of RIL Genotypes	97
4.2. Simulation of Gene Expression Simulation	98
4.3. Multiple Gene/trait JZmapqtl Mapping	99
4.4. Simulation Results	101
5. Discussion	104
6. References	105

Chapter 8

Charting Contig-Component Relationships within the Triticeae: Exploiting the Genome

109

Gerard R. Lazo, Nancy Lui, Frank M. You, David D. Hummel,
Shiaoman Chao, and Olin D. Anderson

1. Introduction	109
2. Approach	110
2.1. Display Settings	111
2.2. Discerning Contig-Component Relationships	112
2.3. Other Applications	115
3. Software	117
4. Discussion	117
5. References	119

Chapter 9

Protein Family Classification with Discriminant Function Analysis ..

121

Etsuko N. Moriyama and Junhyong Kim

1. Introduction	121
1.1. Protein Classification Methods	122
1.2. Pros and Cons for the Current Protein Classification Tools ..	122
1.3. G-protein Coupled Receptor Super Family	124
2. Discriminant Analysis	126
2.1. Input Variables	126

2.2. Datasets	127
2.3. Discrimination Methods	127
2.4. Performance Comparisons for the GPCR Protein Family Classification	128
3. Development of a Hierarchical Classification Algorithm	130
4. References	131

Chapter 10

Exploiting Natural Variation to Understand Gene Function in Pine . 133

David B. Neale and Garth R. Brown

1. Introduction	133
1.1. Association Mapping in Plants	134
2. Pine is a Model Species to Understand Allelic Effects on Phenotypes 134	
2.1. Pines are Evolutionarily Much Older than Crop Species	135
2.2. Pines are Mostly Undomesticated and Are Found in Large Random-Mating Populations	135
2.3. Clonal Replication Permits Precise Evaluation of Phenotypes	136
2.4. Direct Determination of Haplotype	136
3. Association Mapping in Loblolly Pine	138
3.1. Clonal Association Mapping Populations and Phenotypic Evaluations	138
3.2. SNP Discovery, Nucleotide Diversity and Linkage Disequilibrium	138
3.3. Associations between SNP Genotype and Phenotype	142
4. Summary	143
5. References	143

Chapter 11

Merging Analyses of Predisposition and Physiology Towards Polygene Discovery 145

Daniel Pomp, Mark F. Allan, and Stephanie R. Wesolowsk

1. Introduction	145
2. Significance	147
3. Background	147
4. Transcriptome Mapping: A New Experimental Paradigm for Analysis of Complex Trait Genetics	149
4.1. Expected Outcomes from the Transcriptome Mapping Paradigm	151
4.2. Future Directions for Transcriptome Mapping	154
5. Statistical Issues	154

6. Implementation	155
7. References	158

Chapter 12

Mining the EST Databases to Determine Evolutionary Events in the Legumes and Grasses	163
---	------------

Jessica A. Schlueter, Phillip Dixon, Cheryl Granger,
and Randy C. Shoemaker

1. Introduction	163
2. Evolutionary History of Legumes and Grasses	164
3. Identification of Duplicated Contigs	166
4. Pairwise Distance Measures and Mixtures of Normal Distributions	167
5. Coalescence Estimates	169
6. Positively Selected Genes	175
7. Summary	177
8. References	178

Chapter 13

A Biologist's View of Systems Integration Systems Biology: The Pathogen Portal Project	183
---	------------

R. Lathigra, Y. He, R.R. Vines, E.K. Nordberg, and B.W.S. Sobral

1. Introduction	183
2. Illustration of PathPort	185
2.1. Use of PathPort for Genome Annotation and Comparison	185
2.2. PathPort as a Pathogen Information Resource	187
3. Other PathPort Tools for Functional Annotation	191
4. Discussion	192
5. References	195

Chapter 14

Alignment of Wheat and Rice Structural Genomics Resources	197
--	------------

Daryl J. Somers, Sylvie Cloutier, and Travis Banks

1. Introduction	197
2. Primary Structural Genomics Resources	198

2.1. Genetic Maps	198
2.2. QTL Mapping	199
2.3. EST Sequences	199
2.4. Chinese Spring Deletion Lines	199
2.5. Large Insert Libraries	200
2.6. Rice Genome Sequence	201
3. SNPs in Wheat ESTs	202
3.1. Nested PCR Analysis of SNPs	202
3.2. Development of Genome-Specific SNPs	203
4. Wheat/Rice Virtual Mapping (WRVM)	204
5. Implementation of 'WRVM'	207
6. Summary	207
7. References	209

Chapter 15

Computational Identification of Legume-Specific Genes..... 211

Michelle A. Graham, Kevin A. T. Silverstein, Steven B. Cannon,
and Kathryn A. VandenBosch

1. Identification of Sequences of Interest from Public Databases	211
2. Importance of Legumes	212
3. Databases of Expressed Sequence Tags	212
3.1. Data Mining the TIGR Gene Indices	213
3.2. Data Mining the <i>M. truncatula</i> MtDB Database	213
4. Identification of Legume-Specific Genes	214
4.1. Using BLAST to Identify Legume-Specific Genes	214
4.2. Are the Identified Legume-Specific TCs Really Legume-Specific?	217
5. How Can a Gene's Function be Determined When BLAST Fails to Find a Homolog?	218
5.1. Single Linkage Clustering	218
5.2. Motif Searching	220
6. Future Goals	223
7. References	223

ABSTRACTS

Brahms and Beergenes: Information Management for Genetic Research on Barley and Oat 229

Jean Gerster, Nicholas A. Tinker, Yella Jovich-Zahirovich, Anissa Lybaert,
Shaolin Liu, Stephen J. Molnar, and Diane E. Mather

Microcolinearity in the Allotetraploid <i>Gossypium hirsutum</i>	231
C. Grover, H. Kim, R.A. Wing, and J.F. Wendel	
Determining the Chromosomal Location of the Wheat Leaf Rust Resistance Gene <i>LrWI</i>	233
C. Hiebert, J. Thomas, and B. McCallum	
Characterization of EST Libraries from Drought-Stressed Leaves and Penetrated Roots of Rice	235
Md S. Pathan, William G. Spollen, Mark Fredricksen, Hans J. Bohnert, Deshui Zhang, and Henry T. Nguyen	
The Phylogenetic Utility of N-Length DNA Strings in Plants	237
Ryan Rapp and J. Gordon Burleigh	
Telocentrics as a Breeding Tool in Wheat	239
Erica Riedel and Julian Thomas	
Comparative DNA Sequence Analysis of Wheat and Rice Genomes..	241
Mark E. Sorrells and Mauricio La Rota	
Marker Assisted Backcrossing—Some Lessons from Simulation	243
Julian Thomas and Daryl Somers	
<i>Ceregenedb</i>: A Database of Coding Sequence Conservation Between Rice and Nonrice Cereals and <i>Arabidopsis</i>	245
Shibo Zhang, Brian C. Thomas, and Peggy G. Lemaux	
Index	247