
Contents

1	Introduction	1
----------	---------------------	----------

Part I Introduction and Basic Algorithms

2	Basics of Molecular Biology	7
2.1	Proteins	7
2.2	Nucleic Acids	9
2.3	Hereditary Information and Protein Biosynthesis	12
2.4	Experimental Techniques	15
2.4.1	Basic Terms and Methods	15
2.4.2	Duplication of DNA	15
2.4.3	Gel Electrophoresis and Direct Sequencing	16
2.4.4	DNA Chips	19
2.5	Bibliographic Notes	20
3	Basic Concepts: Strings, Graphs, and Algorithms	23
3.1	Strings	23
3.2	Graphs	25
3.3	Algorithms and Complexity	28
3.4	Bibliographic Notes	35
4	String Algorithms	37
4.1	The String Matching Problem	37
4.2	String Matching Automata	39
4.3	The Boyer–Moore Algorithm	44
4.4	Suffix Trees	50
4.5	Further Applications of Suffix Trees	58
4.5.1	Generalized Suffix Trees and the Substring Problem	58
4.5.2	Longest Common Substrings	61
4.5.3	Efficient Computation of Overlaps	63

4.5.4	Repeats in Strings	66
4.6	Suffix Arrays	68
4.7	Summary	77
4.8	Bibliographic Notes	78
5	Alignment Methods	81
5.1	Alignment of Two Strings	82
5.1.1	Basic Definitions	82
5.1.2	Global Alignment	84
5.1.3	Local and Semiglobal Alignment	89
5.1.4	Generalized Scoring Functions	94
5.2	Heuristic Methods for Database Search	97
5.2.1	The FASTA Heuristic	98
5.2.2	The BLAST Heuristic	99
5.3	Multiple Alignments	101
5.3.1	Definition and Scoring of Multiple Alignments	101
5.3.2	Exact Computation of Multiple Alignments	104
5.3.3	Combining Pairwise Alignments	109
5.4	Summary	114
5.5	Bibliographic Notes	114

Part II DNA Sequencing

6	Introduction and Overview	119
7	Physical Mapping	123
7.1	Restriction Site Mapping	123
7.1.1	The Double Digest Approach	124
7.1.2	The Partial Digest Approach	131
7.1.3	Comparison of Methods for Restriction Site Mapping ..	141
7.2	Hybridization Mapping	143
7.2.1	Mapping with Unique Probes	146
7.2.2	Mapping with Unique Probes and Errors	157
7.2.3	Mapping with Non-unique Probes	165
7.3	Summary	166
7.4	Bibliographic Notes	168
8	DNA Sequencing	171
8.1	Shotgun Sequencing	171
8.1.1	Crucial Points to Be Considered in a Suitable Model ..	174
8.1.2	The Shortest Common Superstring Problem	176
8.1.3	Refined Models for Fragment Assembly	196
8.2	Sequencing by Hybridization	201
8.3	Summary	207

8.4 Bibliographic Notes	208
-------------------------------	-----

Part III Analyzing Biological Data

9 Finding Signals in DNA Sequences	213
9.1 Identical and Similar Substrings	213
9.2 Tandem Repeats	217
9.3 Frequent and Infrequent Substrings	223
9.4 Hidden Markov Models.....	228
9.5 Summary.....	235
9.6 Bibliographic Notes	235
10 Genome Rearrangements	237
10.1 Modeling	237
10.2 Sorting Undirected Permutations	239
10.3 Sorting Directed Permutations	247
10.4 Computing the Syntenic Distance	249
10.5 Summary.....	255
10.6 Bibliographic Notes	255
11 Phylogenetic Trees	257
11.1 Ultrametric Distances	258
11.2 Additive Trees	265
11.3 Characters with Binary States	268
11.4 The Parsimony Principle and the Quartet Method	275
11.5 Summary.....	283
11.6 Bibliographic Notes	285
12 Haplotyping	287
12.1 Inferring Haplotypes from a Population	288
12.2 Haplotyping a Single Individual	305
12.3 Summary.....	316
12.4 Bibliographic Notes	316
13 Molecular Structures	319
13.1 RNA Secondary Structure Prediction	320
13.1.1 Minimizing the Free Energy	322
13.1.2 Stochastic Context-Free Grammars	329
13.2 Structure-Based Comparison of Biomolecules.....	337
13.3 Protein Structure Prediction	349
13.3.1 De Novo Structure Prediction — The HP Model	352
13.3.2 Protein Threading.....	363
13.4 Summary.....	369
13.4.1 RNA Secondary Structure Prediction	369

13.4.2 Structure-Based Comparison of Biomolecules	371
13.4.3 Protein Structure Prediction	371
13.5 Bibliographic Notes	372
13.5.1 RNA Secondary Structure Prediction	372
13.5.2 Structure-Based Comparison of Biomolecules	373
13.5.3 Protein Structure Prediction	374
References	377
Index	389



<http://www.springer.com/978-3-540-71912-0>

Algorithmic Aspects of Bioinformatics

Böckenhauer, H.-J.; Bongartz, D.

2007, X, 396 p. 118 illus., Hardcover

ISBN: 978-3-540-71912-0