

# Contents

---

## Part I: Sequence Alignments

---

- 1 Pairwise Sequence Alignment *Benjamin N. Jackson and Srinivas Aluru* 1-1
- 2 Spliced Alignment and Similarity-based Gene Recognition *Alexei D. Neverov, Andrei A. Mironov, and Mikhail S. Gelfand* . . . . . 2-1
- 3 Multiple Sequence Alignment *Osamu Gotoh, Shinsuke Yamada, and Tet-sushi Yada* . . . . . 3-1
- 4 Parametric Sequence Alignment *David Fernández-Baca and Balaji Venkat-achalam* . . . . . 4-1

## Part II: String Data Structures

---

- 5 Lookup Tables, Suffix Trees and Suffix Arrays *Srinivas Aluru and Pang Ko* . . . . . 5-1
- 6 Suffix Tree Applications in Computational Biology *Pang Ko and Srinivas Aluru* . . . . . 6-1
- 7 Enhanced Suffix Arrays and Applications *Mohamed I. Abouelhoda, Stefan Kurtz, and Enno Ohlebusch* . . . . . 7-1

## Part III: Genome Assembly and EST Clustering

---

- 8 Computational Methods for Genome Assembly *Xiaoqiu Huang* . . . . . 8-1
- 9 Assembling the Human Genome *Richa Agarwala* . . . . . 9-1
- 10 Comparative Methods for Sequence Assembly *Vamsi Veeramachaneni* 10-1
- 11 Information Theoretic Approach to Genome Reconstruction *Suchendra Bhandarkar, Jinling Huang, and Jonathan Arnold* . . . . . 11-1
- 12 Expressed Sequence Tags: Clustering and Applications *Anantharaman Kalyanaraman and Srinivas Aluru* . . . . . 12-1
- 13 Algorithms for Large-Scale Clustering and Assembly of Biological Sequence Data *Scott J. Emrich, Anantharaman Kalyanaraman, and Srinivas Aluru* . . . . . 13-1

## Part IV: Genome-Scale Computational Methods

---

- 14 Comparisons of Long Genomic Sequences: Algorithms and Applications *Michael Brudno and Inna Dubchak* . . . . . 14-1
- 15 Chaining Algorithms and Applications in Comparative Genomics *Enno Ohlebusch and Mohamed I. Abouelhoda* . . . . . 15-1
- 16 Computational Analysis of Alternative Splicing *Mikhail S. Gelfand* . . . . . 16-1
- 17 Human Genetic Linkage Analysis *Alejandro A. Schäffer* . . . . . 17-1
- 18 Haplotype Inference *Dan Gusfield and Steven Hecht Orzack* . . . . . 18-1

## Part V: Phylogenetics

---

- 19 An Overview of Phylogeny Reconstruction *C. Randal Linder and Tandy Warnow* . . . . . 19-1
- 20 Consensus Trees and Supertrees *Oliver Eulenstein* . . . . . 20-1
- 21 Large-scale Phylogenetic Analysis *Tandy Warnow* . . . . . 21-1
- 22 High-Performance Phylogeny Reconstruction *David A. Bader and Mi Yan* . . . . . 22-1

## Part VI: Microarrays and Gene Expression Analysis

---

- 23 Microarray Data: Annotation, Storage, Retrieval and Communication *Catherine A. Ball and Gavin Sherlock* . . . . . 23-1
- 24 Computational Methods for Microarray Design *Hui-Hsien Chou* . . . . . 24-1
- 25 Clustering Algorithms for Gene Expression Analysis *Pierre Baldi, G. Wesley Hatfield, and Li M. Fu* . . . . . 25-1
- 26 Biclustering Algorithms: A Survey *Amos Tanay, Roded Sharan, and Ron Shamir* . . . . . 26-1
- 27 Identifying Gene Regulatory Networks from Gene Expression Data *Vladimir Filkov* . . . . . 27-1
- 28 Modeling and Analysis of Gene Networks Using Feedback Control Theory *Hana El Samad and Mustang Rammish* . . . . . 28-1

## Part VII: Computational Structural Biology

---

- 29 Predicting Protein Secondary and Supersecondary Structure *Mona Singh* . . . . . 29-1
- 30 Protein Structure Prediction with Lattice Models *William E. Hart and Alantha Newman* . . . . . 30-1
- 31 Protein Structure Determination via NMR Spectral Data *Guohui Lin, Xin Tu, and Xiang Wan* . . . . . 31-1
- 32 Geometric and Signal Processing of Reconstructed 3D Maps of Molecular Complexes *Chandrajit Bajaj and Zeyun Yu* . . . . . 32-1
- 33 In Search of Remote Homolog *Dong Xu, Ognen Duzlevski, and Xii-Fend Wan* . . . . . 33-1
- 34 Biomolecular Modeling using Parallel Supercomputers *Laxmikant V. Kalé, Klaus Schulten, Robert D. Skeel, Glenn Martyna, Mark Tuckerman, James C. Phillips, Sameer Kumar, and Gengbin Zheng* . . . . . 34-1

## Part VIII: Bioinformatic Databases and Data Mining

---

- 35 String Search in External Memory: Data Structures and Algorithms *Pao-lo Ferragina* . . . . . 35-1
- 36 Index Structures for Approximate Matching in Sequence Databases *Tamer Kahveci and Ambuj K. Singh* . . . . . 36-1
- 37 Algorithms for Motif Search *Sanguthevar Rajasekaran* . . . . . 37-1

38	Data Mining in Computational Biology	Mohammed J. Zaki and Karlton	38-1
	Sequeira		A-1
	Index		

# Sequence Alignments

---

1	Pairwise Sequence Alignment	Benjamin N. Jackson and Srinivas Aravamudan	1-1
	Introduction		
	Global Alignment		
	Dynamic Programming Solution		
	Semiglobal and Local Alignment		
	Space Saving Techniques		
	Heuristic Algorithms		
	Other Gap Penalty Functions		
	Substitution Matrices		
	Local Alignment Database Search		
	Similarity and Distance Measures		
	Normalized Local Alignment		
	Asymptotic Improvements		
	Summary		
2	Spliced Alignment and Similarity-based Gene Recognition	Alexei D. Nayerov, Andrei A. Mironov, and Mikhail B. Gelfand	2-1
	Introduction		
	Spliced Alignment of DNA with ESTs and cDNA		
	Protein-DNA Spliced Alignment		
	Using Local Similarity Identified by BLAST		
	Conclusions		
3	Multiple Sequence Alignment	Osamu Gotoh, Shinzuke Yamada, and Tetsushi Yada	3-1
	Biological Background		
	Definitions and Properties		
	Algorithms		
	Methods for Assessment		
	Summary		
4	Parametric Sequence Alignment	David Fernández-Bacá and Balaji Venkateshbabu	4-1
	Introduction		
	Formulation		
	Combinatorial Complexity		
	Parametric Search		
	Constructing the Parameter Space Decomposition		
	Parametric Problems in Hidden Markov Models of Sequence Alignment		
	Notes and Comments		