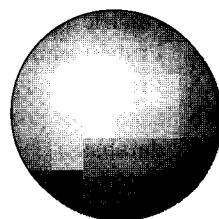


Full Contents



Preface

1 INTRODUCTION: THE REVOLUTION IN BIOLOGICAL INFORMATION

- 1.1 Data explosions
- 1.2 Genomics and high-throughput techniques
- 1.3 What is bioinformatics?
- 1.4 The relationship between population genetics, molecular evolution, and bioinformatics

Summary • References • Problems

2 NUCLEIC ACIDS, PROTEINS, AND AMINO ACIDS

- 2.1 Nucleic acid structure
- 2.2 Protein structure
- 2.3 The central dogma
- 2.4 Physico-chemical properties of the amino acids and their importance in protein folding

Box 2.1 Polymerase chain reaction (PCR)

- 2.5 Visualization of amino acid properties using principal component analysis

- 2.6 Clustering amino acids according to their properties

Box 2.2 Principal component analysis in more detail

Summary • References • Self-test

Biological background

x 3 MOLECULAR EVOLUTION AND POPULATION GENETICS

- 3.1 What is evolution? 37
- 3.2 Mutations 39
- 3.3 Sequence variation within and between species 40
- 3.4 Genealogical trees and coalescence 44
- 3.5 The spread of new mutations 46
- 3.6 Neutral evolution and adaptation 49

Box 3.1 The influence of selection on the fixation probability 50

Box 3.2 A deterministic theory for the spread of mutations 51

Summary • References • Problems 54

4 MODELS OF SEQUENCE EVOLUTION

- 4.1 Models of nucleic acid sequence evolution 58

Box 4.1 Solution of the Jukes–Cantor model 61

4.2 The PAM model of protein sequence evolution 65

Box 4.2 PAM distances 70

4.3 Log-odds scoring matrices for amino acids 71

Summary • References • Problems • Self-test Molecular evolution 76

5 INFORMATION RESOURCES FOR GENES AND PROTEINS

- 5.1 Why build a database? 81
- 5.2 Database file formats 82

5.3	Nucleic acid sequence databases	83	8.6	The maximum-likelihood criterion	173
5.4	Protein sequence databases	89	 8.2	Calculating the likelihood of the data on a given tree	174
5.5	Protein family databases	95	8.7	The parsimony criterion	177
5.6	Composite protein pattern databases	108	8.8	Other methods related to maximum likelihood	179
5.7	Protein structure databases	111	 8.3	Calculating posterior probabilities	182
5.8	Other types of biological database	113	Summary • References • Problems •		
Summary • References		115	Self-test Phylogenetic methods		185
6 SEQUENCE ALIGNMENT ALGORITHMS		119	9 PATTERNS IN PROTEIN FAMILIES		195
6.1	What is an algorithm?	119	9.1	Going beyond pairwise alignment methods for database searches	195
6.2	Pairwise sequence alignment – The problem	121	9.2	Regular expressions	197
6.3	Pairwise sequence alignment – Dynamic programming methods	123	9.3	Fingerprints	200
6.4	The effect of scoring parameters on the alignment	127	9.4	Profiles and PSSMs	205
6.5	Multiple sequence alignment	130	9.5	Biological applications – G protein-coupled receptors	208
Summary • References • Problems		136	Summary • References • Problems •		
7 SEARCHING SEQUENCE DATABASES		139	Self-test Protein families and databases		216
7.1	Similarity search tools	139	10 PROBABILISTIC METHODS AND MACHINE LEARNING		227
7.2	Alignment statistics (in theory)	147	10.1	Using machine learning for pattern recognition in bioinformatics	227
 7.1	Extreme value distributions	151	10.2	Probabilistic models of sequences – Basic ingredients	228
 7.2	Derivation of the extreme value distribution in the word-matching example	152	 10.1	Dirichlet prior distributions	232
7.3	Alignment statistics (in practice)	153	10.3	Introducing hidden Markov models	234
Summary • References • Problems •			 10.2	The Viterbi algorithm	238
Self-test Alignments and database searching		155	 10.3	The forward and backward algorithms	239
8 PHYLOGENETIC METHODS		158	10.4	Profile hidden Markov models	241
8.1	Understanding phylogenetic trees	158	10.5	Neural networks	244
8.2	Choosing sequences	161	 10.4	The back-propagation algorithm	249
8.3	Distance matrices and clustering methods	162	10.6	Neural networks and protein secondary structure prediction	250
 8.1	Calculation of distances in the neighbor-joining method	167	Summary • References • Problems		253
8.4	Bootstrapping	169			
8.5	Tree optimization criteria and tree search methods	171			

11 FURTHER TOPICS IN MOLECULAR EVOLUTION AND PHYLOGENETICS

11.1 RNA structure and evolution	257
11.2 Fitting evolutionary models to sequence data	257
11.3 Applications of molecular phylogenetics	266
Summary • References	272
	279

12 GENOME EVOLUTION

12.1 Prokaryotic genomes	283
 12.1 Web resources for bacterial genomes	283
12.2 Organellar genomes	284
Summary • References	298
	309

13 DNA MICROARRAYS AND THE 'OMES

13.1 'Omes and 'omics	313
13.2 How do microarrays work?	313
13.3 Normalization of microarray data	314
13.4 Patterns in microarray data	316
13.5 Proteomics	319
13.6 Information management for the 'omes	325
	330

13.1 Examples from the Gene Ontology

Summary • References • Self-test	335
	337

MATHEMATICAL APPENDIX

M.1 Exponentials and logarithms	343
M.2 Factorials	343
M.3 Summations	344
M.4 Products	344
M.5 Permutations and combinations	345
M.6 Differentiation	345
M.7 Integration	346
M.8 Differential equations	347
M.9 Binomial distributions	347
M.10 Normal distributions	348
M.11 Poisson distributions	348
M.12 Chi-squared distributions	350
M.13 Gamma functions and gamma distributions	351
Problems • Self-test	352
	353

List of Web addresses	355
-----------------------	-----

Glossary	357
----------	-----

Index	363
-------	-----