Bioinformatics and Molecular Evolution

Paul G. Higgs and Teresa K. Attwood
Preface x
Chapter plan xiii
1 Introduction: The revolution in biological information 1
2 Nucleic acids, proteins, and amino acids 12
3 Molecular evolution and population genetics 37
4 Models of sequence evolution 58
5 Information resources for genes and proteins 81
6 Sequence alignment algorithms 119
7 Searching sequence databases 139
8 Phylogenetic methods 158
9 Patterns in protein families 195
10 Probabilistic methods and machine learning 227
11 Further topics in molecular evolution and phylogenetics 257
12 Genome evolution 283
13 DNA Microarrays and the 'omes 313
Mathematical appendix 343
List of Web addresses 355
Glossary 357
Index 363

Color plates fall between pages 194 and 195
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

3 MOLECULAR EVOLUTION AND POPULATION GENETICS
3.1 What is evolution?
3.2 Mutations
3.3 Sequence variation within and between species
3.4 Genealogical trees and coalescence
3.5 The spread of new mutations
3.6 Neutral evolution and adaptation
Box 3.1 The influence of selection on the fixation probability
Box 3.2 A deterministic theory for the spread of mutations
Summary • References • Problems

4 MODELS OF SEQUENCE EVOLUTION
4.1 Models of nucleic acid sequence evolution
Box 4.1 Solution of the Jukes–Cantor model
4.2 The PAM model of protein sequence evolution
Box 4.2 PAM distances
4.3 Log-odds scoring matrices for amino acids
Summary • References • Problems • Self-test

5 INFORMATION RESOURCES FOR GENES AND PROTEINS
5.1 Why build a database?
5.2 Database file formats

Summary • References • Problems • Self-test

Biological background
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  272
References  279

12 GENOME EVOLUTION
12.1 Prokaryotic genomes 283
12.2 Organellar genomes 298
Summary  309

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

10  335
11  337
12  343
13  355

Box 13.1 Examples from the Gene Ontology
Summary  335
References  337

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

List of Web addresses  355
Glossary  357
Index  363