

Contents

	<i>page</i>
<i>Preface</i>	<i>ix</i>
1 Introduction	1
1.1 <i>Sequence similarity, homology, and alignment</i>	2
1.2 <i>Overview of the book</i>	2
1.3 <i>Probabilities and probabilistic models</i>	4
1.4 <i>Further reading</i>	10
2 Pairwise alignment	12
2.1 <i>Introduction</i>	12
2.2 <i>The scoring model</i>	13
2.3 <i>Alignment algorithms</i>	17
2.4 <i>Dynamic programming with more complex models</i>	28
2.5 <i>Heuristic alignment algorithms</i>	32
2.6 <i>Linear space alignments</i>	34
2.7 <i>Significance of scores</i>	36
2.8 <i>Deriving score parameters from alignment data</i>	41
2.9 <i>Further reading</i>	45
3 Markov chains and hidden Markov models	46
3.1 <i>Markov chains</i>	48
3.2 <i>Hidden Markov models</i>	51
3.3 <i>Parameter estimation for HMMs</i>	62
3.4 <i>HMM model structure</i>	68
3.5 <i>More complex Markov chains</i>	72
3.6 <i>Numerical stability of HMM algorithms</i>	77
3.7 <i>Further reading</i>	79
4 Pairwise alignment using HMMs	80
4.1 <i>Pair HMMs</i>	81
4.2 <i>The full probability of x and y, summing over all paths</i>	87
4.3 <i>Suboptimal alignment</i>	89
4.4 <i>The posterior probability that x_i is aligned to y_j</i>	91
4.5 <i>Pair HMMs versus FSAs for searching</i>	95

vi	<i>Contents</i>	
4.6	<i>Further reading</i>	98
5	Profile HMMs for sequence families	100
5.1	<i>Ungapped score matrices</i>	102
5.2	<i>Adding insert and delete states to obtain profile HMMs</i>	102
5.3	<i>Deriving profile HMMs from multiple alignments</i>	105
5.4	<i>Searching with profile HMMs</i>	108
5.5	<i>Profile HMM variants for non-global alignments</i>	113
5.6	<i>More on estimation of probabilities</i>	115
5.7	<i>Optimal model construction</i>	122
5.8	<i>Weighting training sequences</i>	124
5.9	<i>Further reading</i>	132
6	Multiple sequence alignment methods	134
6.1	<i>What a multiple alignment means</i>	135
6.2	<i>Scoring a multiple alignment</i>	137
6.3	<i>Multidimensional dynamic programming</i>	141
6.4	<i>Progressive alignment methods</i>	143
6.5	<i>Multiple alignment by profile HMM training</i>	149
6.6	<i>Further reading</i>	159
7	Building phylogenetic trees	160
7.1	<i>The tree of life</i>	160
7.2	<i>Background on trees</i>	161
7.3	<i>Making a tree from pairwise distances</i>	165
7.4	<i>Parsimony</i>	173
7.5	<i>Assessing the trees: the bootstrap</i>	179
7.6	<i>Simultaneous alignment and phylogeny</i>	180
7.7	<i>Further reading</i>	188
7.8	<i>Appendix: proof of neighbour-joining theorem</i>	190
8	Probabilistic approaches to phylogeny	192
8.1	<i>Introduction</i>	192
8.2	<i>Probabilistic models of evolution</i>	193
8.3	<i>Calculating the likelihood for ungapped alignments</i>	197
8.4	<i>Using the likelihood for inference</i>	205
8.5	<i>Towards more realistic evolutionary models</i>	215
8.6	<i>Comparison of probabilistic and non-probabilistic methods</i>	224
8.7	<i>Further reading</i>	231
9	Transformational grammars	233
9.1	<i>Transformational grammars</i>	234
9.2	<i>Regular grammars</i>	237
9.3	<i>Context-free grammars</i>	242

Cambridge University Press

978-0-521-62971-3 - Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids

Richard Durbin, Sean R. Eddy, Anders Krogh and Graeme Mitchison

Table of Contents

[More information](#)

<i>Contents</i>		vii
9.4	<i>Context-sensitive grammars</i>	247
9.5	<i>Stochastic grammars</i>	250
9.6	<i>Stochastic context-free grammars for sequence modelling</i>	252
9.7	<i>Further reading</i>	259
10	RNA structure analysis	260
10.1	<i>RNA</i>	261
10.2	<i>RNA secondary structure prediction</i>	267
10.3	<i>Covariance models: SCFG-based RNA profiles</i>	277
10.4	<i>Further reading</i>	297
11	Background on probability	299
11.1	<i>Probability distributions</i>	299
11.2	<i>Entropy</i>	305
11.3	<i>Inference</i>	311
11.4	<i>Sampling</i>	314
11.5	<i>Estimation of probabilities from counts</i>	319
11.6	<i>The EM algorithm</i>	323
	<i>Bibliography</i>	326
	<i>Author index</i>	345
	<i>Subject index</i>	350