

Contents

ACKNOWLEDGEMENTS	XI
PREFACE	XIII
1. INSTALLATION OF DAMBE AND A QUICK START.....	1
1. INSTALLATION.....	1
2. A JUMP START	2
2. FILE CONVERSION.....	7
1. A PLETHORA OF COMPUTER PROGRAMS	8
2. A PLETHORA OF SEQUENCE FORMATS	8
3. READSEQ.....	9
4. FILE CONVERSION USING DAMBE.....	10
4.1 <i>Convert all sequences from one format to another</i>	11
4.2 <i>Converting a subset of sequences</i>	12
4.3 <i>Output PHYLTEST files</i>	13
3. PROCESSING GENBANK FILES.....	17
1. GENBANK FILE FORMAT	18
2. READING GENBANK FILES WITH DAMBE.....	20
4. ACCESSING GENBANK OR NETWORKED COMPUTERS	25
1. INTRODUCTION.....	25
2. READING MOLECULAR SEQUENCES DIRECTLY FROM GENBANK.....	25
3. READING FROM AND WRITING TO ANOTHER NETWORKED COMPUTER	30

4. EXERCISE	32
5. PAIR-WISE AND MULTIPLE SEQUENCE ALIGNMENT.....	33
1. INTRODUCTION.....	33
1.1 <i>The dot-matrix approach</i>	33
1.2 <i>Similarity or distance method</i>	36
2. SEQUENCE ALIGNMENT USING DAMBE.....	37
2.1 <i>Align nucleotide or amino acid sequences</i>	37
2.2 <i>Align nucleotide sequences against amino acid sequences</i>	38
6. FACTORS AFFECTING NUCLEOTIDE FREQUENCIES.....	41
1. INTRODUCTION.....	41
1.1 <i>The frequency parameters</i>	41
1.2 <i>Factors that might change the frequency parameters</i>	42
1.3 <i>Frequency parameters and phylogenetic analyses</i>	43
2. COUNTING NUCLEOTIDE AND DINUCLEOTIDE FREQUENCIES.....	44
7. CASE STUDY 1: ARTHROPOD PHYLOGENY.....	49
1. INTRODUCTION.....	49
2. OBTAIN DATA FROM GENBANK.....	50
3. ALIGN THE SEQUENCES.....	53
4. DATA ANALYSIS	56
8. FACTORS AFFECTING CODON FREQUENCIES.....	59
1. INTRODUCTION.....	59
2. GENERATING CODON USAGE TABLE WITH DAMBE.....	60
3. DNA METHYLATION AND USAGE OF ARGININE CODONS.....	64
4. TRANSCRIPTION EFFICIENCY AND CODON USAGE BIAS	66
5. TRANSLATIONAL EFFICIENCY AND CODON USAGE BIAS	66
6. CODON FREQUENCY AND PEPTIDE LENGTH IN ANCIENT PROTEINS	68
9. CASE STUDY 2: TRANSCRIPTION AND CODON USAGE BIAS.....	71
1. INTRODUCTION.....	71
2. MAXIMIZING TRANSCRIPTIONAL EFFICIENCY.....	72
3. PREDICTIONS AND EMPIRICAL TESTS	75
4. AN ALTERNATIVE EXPLANATION.....	85
5. DISCUSSION.....	89
10. CASE STUDY 3: TRANSLATION AND CODON USAGE BIAS.....	91
1. INTRODUCTION.....	91
2. THE ELONGATION MODEL, ITS PREDICTIONS, AND EMPIRICAL TESTS	92
2.1 <i>Adaptation of Codon Usage to tRNA Content</i>	94
2.2 <i>Adaptation of tRNA to Codon Usage</i>	98

2.3	<i>Evolution of tRNA in Response to Amino Acid Usage</i>	99
2.4	<i>Translational Efficiency and Translational Accuracy</i>	102
3.	DISCUSSION.....	103
3.1	<i>Validity of the Model</i>	103
3.2	<i>Translational Efficiency and Accuracy on Codon Usage Bias</i>	104
3.3	<i>How Optimized Are the Translational Machinery?</i>	105
11.	EVOLUTION OF AMINO ACID USAGE	107
1.	INTRODUCTION.....	107
2.	AMINO ACID USAGE BIAS.....	109
12.	PATTERN OF NUCLEOTIDE SUBSTITUTIONS	115
1.	INTRODUCTION.....	115
2.	USE DAMBE TO DOCUMENT EMPIRICAL SUBSTITUTION PATTERNS	118
2.1	<i>Simple output</i>	118
2.2	<i>Detailed Output</i>	119
13.	PREAMBLE TO THE PATTERN OF CODON SUBSTITUTION	125
1.	INTRODUCTION.....	125
2.	DEFAULT SUBSTITUTION PATTERNS WITH NO SELECTION	126
14.	FACTORS AFFECTING CODON SUBSTITUTIONS	131
1.	INTRODUCTION.....	131
1.1	<i>The Rate of Codon Substitutions and its Determinants</i>	131
1.2	<i>Models of Codon Substitution</i>	132
1.3	<i>The Expected Pattern of Nonsynonymous Codon Substitutions</i>	134
2.	CODON COMPARISON WITH DAMBE	136
2.1	<i>Tracing evolutionary history</i>	136
2.2	<i>Summary of codon substitution pattern</i>	140
2.3	<i>Single-step Nonsynonymous Codon Substitutions</i>	142
15.	CASE STUDY 4: TRANSITION BIAS	147
1.	INTRODUCTION.....	147
2.	GET SEQUENCE DATA	151
3.	DATA ANALYSIS	152
3.1	<i>Phylogeny reconstruction</i>	152
3.2	<i>Pair-wise comparisons between neighboring nodes</i>	157
4.	RESULTS.....	160
5.	DISCUSSION.....	162
16.	SUBSTITUTION PATTERN IN AMINO ACID SEQUENCES	165
1.	SUBSTITUTION PATTERN FROM SEQUENCES IN RST FORMAT.....	165
2.	SUBSTITUTION PATTERN FROM ALL PAIR-WISE COMPARISONS.....	169

17.	A STATISTICAL DIGRESSION.....	171
1.	INTRODUCTION.....	171
2.	TWO DISCRETE PROBABILITY DISTRIBUTIONS.....	172
2.1	<i>The Binomial Distribution and the Goodness-of-fit test</i>	172
2.2	<i>The Multinomial Distribution</i>	174
3.	THE SIMPLEST PRESENTATION OF THE MAXIMUM LIKELIHOOD METHOD..	175
4.	BIAS IN THE MAXIMUM LIKELIHOOD METHOD	177
5.	EXERCISE.....	178
18.	THEORETICAL BACKGROUND OF GENETIC DISTANCES	179
1.	INTRODUCTION.....	179
2.	GENETIC DISTANCES FROM NUCLEOTIDE SEQUENCES	180
2.1	<i>JC69 and TN84 distances</i>	181
2.2	<i>Kimura's two parameter distance</i>	183
2.3	<i>F84 distance</i>	184
2.4	<i>TN93 distance</i>	185
2.5	<i>Lake's paralinear distance</i>	186
3.	DISTANCES BASED ON CODON SEQUENCES	187
3.1	<i>The empirical counting approach</i>	188
3.2	<i>Codon-based maximum likelihood method</i>	190
4.	DISTANCES BASED ON AMINO ACID SEQUENCES	192
5.	GENETIC DISTANCES FROM ALLELE FREQUENCIES	193
5.1	<i>Net's genetic distance</i>	194
5.2	<i>Cavalli-Sforza's chord measure</i>	195
5.3	<i>Reynolds, Weir, and Cockerham's genetic distance</i>	196
19.	MOLECULAR PHYLOGENETICS: CONCEPTS AND PRACTICE	197
1.	THE MOLECULAR CLOCK AND ITS CALIBRATION.....	198
1.1	<i>Calibrating a molecular clock</i>	200
1.2	<i>Complications in calibrating a molecular clock</i>	201
2.	COMMON APPROACHES IN MOLECULAR PHYLOGENETICS	204
2.1	<i>Distance methods</i>	204
2.2	<i>Maximum parsimony method</i>	214
2.3	<i>Maximum likelihood method</i>	216
2.4	<i>Reconstructing Ancestral Sequences</i>	221
3.	EXERCISE	224
20.	TESTING THE MOLECULAR CLOCK HYPOTHESIS.....	225
1.	THE T-TEST.....	226
2.	THE LIKELIHOOD RATIO TEST	227
3.	TEST THE MOLECULAR CLOCK HYPOTHESIS.....	230
21.	TESTING PHYLOGENETIC HYPOTHESES.....	233

1. BASIC STATISTICAL CONCEPTS	234
2. TESTING PHYLOGENETIC HYPOTHESES WITH THE DISTANCE METHOD	236
2.1 <i>The Rationale</i>	236
2.2 <i>Test alternative phylogenetic hypotheses with the distance method</i> ..	238
3. TESTING PHYLOGENETIC HYPOTHESES WITH THE PARSIMONY METHOD ..	241
4. TESTING PHYLOGENETIC HYPOTHESES WITH THE LIKELIHOOD METHOD..	243
5. RESAMPLING METHODS	247
6. EXERCISE.....	248
22. FITTING PROBABILITY DISTRIBUTIONS	249
1. INTRODUCTION.....	249
1.1 <i>The Poisson distribution</i>	250
1.2 <i>The negative binomial distribution</i>	252
1.3 <i>The gamma distribution</i>	254
1.4 <i>Some general guidelines for fitting statistical distributions</i>	257
2. FITTING DISCRETE DISTRIBUTIONS WITH DAMBE.....	258
3. ESTIMATING THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION	261
4. EXERCISE.....	263
LITERATURE CITED.....	265
INDEX	275