

**Lecture Notes:**  
**The Mathematics of Phylogenetics**

Elizabeth S. Allman,  
John A. Rhodes

IAS/Park City Mathematics Institute  
June-July, 2005

Univeristy of Alaska Fairbanks  
Spring 2009, 2012



# Contents

<b>1</b>	<b>Molecular Evolution</b>	<b>1</b>
1.1	DNA structure . . . . .	2
1.2	Mutations . . . . .	3
1.3	Aligned Orthologous Sequences . . . . .	4
<b>2</b>	<b>Combinatorics of Trees I</b>	<b>5</b>
2.1	Graphs and Trees . . . . .	5
2.2	Counting Binary Trees . . . . .	8
2.3	Metric Trees . . . . .	10
2.4	Molecular Clock Trees . . . . .	11
2.5	Rooting Trees with Outgroups . . . . .	12
2.6	Exercises . . . . .	13
<b>3</b>	<b>Parsimony</b>	<b>17</b>
3.1	The Parsimony Criterion . . . . .	17
3.2	The Fitch-Hartigan Algorithm . . . . .	20
3.3	Informative Characters . . . . .	25
3.4	Complexity . . . . .	25
3.5	Weighted Parsimony . . . . .	26
3.6	Recovering Minimal Extensions . . . . .	28
3.7	Further Issues . . . . .	29
3.8	Exercises . . . . .	30
<b>4</b>	<b>Combinatorics of Trees II</b>	<b>35</b>
4.1	Splits . . . . .	35
4.2	Refinements and Consensus Trees . . . . .	38
4.3	Quartets . . . . .	39
4.4	Supertrees . . . . .	40
4.5	Exercises . . . . .	41
<b>5</b>	<b>Distance Algorithms</b>	<b>43</b>
5.1	UPGMA . . . . .	44
5.2	Unequal Branch Lengths . . . . .	47
5.3	The Four-point Condition . . . . .	51

5.4	The Neighbor Joining Algorithm . . . . .	54
5.5	Exercises . . . . .	56
<b>6</b>	<b>Probabilistic Models of DNA Mutation</b>	<b>65</b>
6.1	A simple example . . . . .	65
6.2	Markov Models of DNA Base Substitution on Trees . . . . .	70
6.3	Jukes-Cantor and Kimura Models . . . . .	76
6.4	Time-reversible Models . . . . .	80
6.5	Exercises . . . . .	82
<b>7</b>	<b>Phylogenetic Distances</b>	<b>89</b>
7.1	Jukes-Cantor Distance . . . . .	89
7.2	Kimura Distances . . . . .	92
7.3	Log-det Distance . . . . .	93
7.4	Exercises . . . . .	94
<b>8</b>	<b>Maximum Likelihood</b>	<b>99</b>
8.1	Probabilities and Likelihoods . . . . .	99
8.2	ML Estimators for One-edge Trees . . . . .	103
8.3	Inferring Trees by ML . . . . .	105
8.4	Efficient ML Computation . . . . .	106
8.5	Exercises . . . . .	109
<b>9</b>	<b>Tree Space</b>	<b>111</b>
9.1	What is Tree Space? . . . . .	111
9.2	Moves in Tree Space . . . . .	112
9.3	Searching Tree space . . . . .	118
9.4	Metrics on Tree Space . . . . .	120
9.5	Additional Remarks . . . . .	122
9.6	Exercises . . . . .	122
<b>10</b>	<b>Rate-variation and Mixture Models</b>	<b>125</b>
10.1	Invariable Sites Models . . . . .	125
10.2	Rates-Across-Sites Models . . . . .	127
10.3	The Covarion Model . . . . .	130
10.4	Exercises . . . . .	133
<b>11</b>	<b>Consistency and Long Branch Attraction</b>	<b>135</b>
11.1	Statistical Consistency . . . . .	136
11.2	Parsimony and Consistency . . . . .	137
11.3	Consistency of Distance Methods . . . . .	140
11.4	Consistency of Maximum Likelihood . . . . .	141
11.5	Performance on Short Sequences . . . . .	142
11.6	Exercises . . . . .	143
<b>12</b>	<b>Bayesian Inference</b>	<b>145</b>

<i>CONTENTS</i>	v
<b>13 Gene trees and species trees</b>	<b>147</b>
<b>Bibliography</b>	<b>149</b>

