
Contents

Preface	V
1 Basic Models	1
1.1 ATGCs of life	1
1.2 Wright-Fisher model	5
1.2.1 The coalescent	8
1.2.2 Shape of the genealogical tree	11
1.3 Infinite alleles model	14
1.3.1 Hoppe's urn, Ewens' sampling formula	15
1.3.2 Chinese restaurants and sufficient statistics	20
1.3.3 Branching process viewpoint	26
1.4 Infinite sites model	29
1.4.1 Segregating sites	31
1.4.2 Nucleotide diversity	39
1.4.3 Pairwise differences	40
1.4.4 Folded site frequency spectrum	43
1.5 Moran model	46
1.5.1 Fixation probability and time	47
1.5.2 Site frequency spectrum mean	50
2 Estimation and Hypothesis Testing	53
2.1 Site frequency spectrum covariance	53
2.2 Estimates of θ	59
2.3 Hypothesis testing overview	63
2.4 Difference statistics	65
2.4.1 Tajima's D	65
2.4.2 Fu and Li's D	67
2.4.3 Fay and Wu's H	69
2.4.4 Conditioning on S_n	71
2.5 The HKA test	72
2.6 McDonald-Kreitman test	78

3	Recombination	83
3.1	Two loci	83
3.1.1	Sample of size 2	83
3.1.2	Sample of size n	87
3.2	m loci	90
3.2.1	Samples of size 2	91
3.2.2	Samples of size n	91
3.2.3	Pairwise differences	93
3.3	Linkage disequilibrium	97
3.4	Ancestral recombination graph	101
3.4.1	Simulation	105
3.4.2	Two approximate algorithms	107
3.5	Counting recombinations	111
3.6	Estimating recombination rates	114
3.6.1	Equations for the two-locus sampling distribution	115
3.6.2	Simulation methods	118
3.6.3	Composite likelihood estimation of ρ	120
3.7	Haplotypes and hot spots	122
4	Population Complications	125
4.1	Large family sizes	125
4.2	Population growth	130
4.2.1	Exponential growth	131
4.2.2	Sudden population expansion	134
4.3	Founding effects and bottlenecks	138
4.4	Effective population size	143
4.5	Matrix migration models	146
4.5.1	Strobeck's theorem	148
4.5.2	Fast migration limit	149
4.6	Symmetric island model	150
4.6.1	Identity by descent	150
4.6.2	Mean and variance of coalescence times	152
4.6.3	Effective population sizes	153
4.6.4	Large n limit	154
4.7	Fixation indices	157
5	Stepping Stone Model	161
5.1	$d = 1$, Exact results	161
5.2	$d = 1$ and 2, Fourier methods	165
5.3	$d = 2$, Coalescence times	172
5.3.1	Random walk results	172
5.3.2	Samples of size 2	175
5.3.3	Fixation indices F_{ST}	177
5.4	$d = 2$, Genealogies	179
5.4.1	Simulation results	181

5.5	$d = 1$, Continuous models	183
5.6	$d = 2$, Continuous models	188
6	Natural Selection	191
6.1	Directional selection	191
6.1.1	Fixation probability	192
6.1.2	Time to fixation	194
6.1.3	Three phases of the fixation process	197
6.1.4	Ancestral selection graph	199
6.2	Balancing selection	201
6.3	Background selection	211
6.4	Muller's ratchet	219
6.4.1	Evolutionary advantages of recombination	220
6.4.2	Sex, epistasis, and Kondrashov	222
6.5	Hitchhiking	225
6.6	Better approximations	234
6.7	Recurrent sweeps	241
6.7.1	Nucleotide diversity	241
6.7.2	Genealogies	243
6.7.3	Segregating sites	245
7	Diffusion Processes	249
7.1	Infinitesimal mean and variance	250
7.2	Examples of diffusions	252
7.3	Transition probabilities	258
7.4	Hitting probabilities	262
7.5	Stationary measures	268
7.6	Occupation times	273
7.7	Green's functions	276
7.8	Examples	280
7.9	Conditioned processes	287
7.10	Boundary behavior	293
7.11	Site frequency spectrum	298
7.11.1	Poisson random field model	300
7.12	Fluctuating selection	304
8	Multidimensional Diffusions	313
8.1	K allele model	313
8.1.1	Fixation probabilities and time	315
8.1.2	Stationary distributions	318
8.2	Recombination	322
8.2.1	A clever change of variables	323
8.2.2	Time-dependent behavior	326
8.2.3	Equilibrium when there is mutation	329
8.3	Hill-Robertson interference	334

8.4	Gene duplication	341
8.5	Watterson's double recessive null model	343
8.6	Subfunctionalization	347
9	Genome Rearrangement	355
9.1	Inversions	355
9.1.1	Breakpoint graph.....	355
9.1.2	Hurdles	358
9.2	When is parsimony reliable?	362
9.2.1	Phase transition.....	365
9.2.2	Bayesian approach.....	367
9.3	Nadeau and Taylor's analysis	370
9.4	Genomic distance.....	375
9.4.1	Graph distance.....	375
9.4.2	Bayesian estimation	377
9.5	Midpoint problem	382
9.6	Genome duplication	389
9.6.1	Yeast	389
9.6.2	Maize	393
9.6.3	Arabidopsis thaliana	396
	References	399
	Index	427



<http://www.springer.com/978-0-387-78168-6>

Probability Models for DNA Sequence Evolution

Durrett, R.

2008, XII, 431 p., Hardcover

ISBN: 978-0-387-78168-6