CONTENTS

PREFACE XV

CHAPTER 1
MODELS OF EVOLUTION 1

BASIC SEQUENCE TERMINOLOGY 2

MODELS OF EVOLUTIONARY PROCESSES 3
Models of populations 3
The effective size of a population 6
Models of mutation 7
Models of recombination 11
Models of natural selection 12
Models of migration 14
Models of mating 17

MODELS OF MOLECULAR EVOLUTION 18
The neutral theory of molecular evolution 18
Misunderstanding and misuse of the neutral theory 21
Alternative models of molecular evolution 23

CHAPTER 2
EXPERIMENTAL DESIGN 25

POPULATION SAMPLING 25

DNA SEQUENCING 26

Sanger sequencing 27

Haplotypes 30

Next-generation sequencing 31

The effect of short sequence reads 33

The effect of high error rates 35

Diverse experimental designs for next-generation genome sequencing 36

GENOTYPING 40

CHAPTER 3
DESCRIBING VARIATION 43

MEASURES OF SEQUENCE DIVERSITY 44

Common estimators of θ for SNPs under the infinite sites model 44

The frequency spectrum of alleles 47

Estimators of θ for SNPs under finite sites models and incorporating sequencing error 50

Interpreting measures of nucleotide variation 53

ADDITIONAL MEASURES OF DIVERSITY 55

Microsatellite variation 55

Haplotype variation 57

CHAPTER 4
RECOMBINATION 59

MEASURING LINKAGE DISEQUILIBRIUM 59

Linkage and linkage disequilibrium 59

Measures of pairwise linkage disequilibrium 61

Summarizing pairwise linkage disequilibrium 63
Extensions to multiple alleles, multiple loci, and genotypes 66

ESTIMATING RECOMBINATION 68
Finding the minimum number of recombination events in a sample 68
The population recombination parameter 70
Estimating the population recombination parameter 71

GENE CONVERSION 77

CHAPTER 5
POPULATION STRUCTURE 79

POPULATION DIFFERENTIATION 79
Populations, subpopulations, and demes 79
The Wahlund effect 80

MEASURING POPULATION DIFFERENTIATION 81
Measuring population differentiation using F_{ST} 81
Alternative measures of population differentiation 86
Is there evidence for population differentiation at a locus? 89

THE EFFECT OF EVOLUTIONARY PROCESSES ON DIFFERENTIATION 91
The effects of natural selection on population differentiation 91
The effects of migration and drift on population differentiation 93
Distinguishing between migration and drift: The isolation-with-migration model 97
Additional methods for detecting migration between pairs of populations 100

DEFINING POPULATIONS 104
Identifying populations and the individuals within them 104
Using the Wahlund effect to identify population structure 105
Evolutionary inference from the determination of population structure 107

CHAPTER 6
THE COALESCENT 111

SIMULATING SAMPLES OF DNA SEQUENCES 111
SIMULATING THE COALESCENT 114
Coalescent genealogies 114
Genealogies and phylogenies 118
UNDERSTANDING THE COALESCENT 118
Important genealogical quantities 118
The coalescent and measures of polymorphism 120
The effects of ascertainment bias on the allele frequency spectrum 122
EXTENDING THE COALESCENT 124
The coalescent with recombination 124
The coalescent and reference genomes 128

CHAPTER 7
DIRECT SELECTION 131

THE ACCUMULATION OF SEQUENCE DIVERGENCE 131
The rate of nucleotide substitution 131
Estimating sequence divergence 134
DETECTING SELECTION USING DIVERGENCE 137
d_N, d_S, and d_N/d_S 137
Interpreting d_N/d_S 141
DETECTING SELECTION USING POLYMORPHISM 145
The effect of selection on the frequency of polymorphism 145
π_N/π_S 148
DETECTING SELECTION USING POLYMORPHISM AND DIVERGENCE 149
The McDonald-Kreitman test 149
Experimental design for the McDonald-Kreitman test 149

Expectations for levels of polymorphism and divergence 151
Interpreting the McDonald-Kreitman test 152
Methodological extensions of the MK test 154
Assumptions of the MK test 156
Biological extensions of the MK test 159
Summarizing selection across the genome 160

CHAPTER 8
LINKED SELECTION 165

DETECTING SELECTION USING THE AMOUNT OF POLYMORPHISM 165
The effects of positive selection on levels of linked neutral variation 165
The effects of balancing selection and strongly deleterious alleles on levels of linked neutral variation 168
The most underappreciated result in molecular population genetics 170
Detecting selection via levels of polymorphism: The HKA test 171
Experimental design for the HKA test 172
Interpreting the HKA test 173

DETECTING SELECTION USING THE ALLELE FREQUENCY SPECTRUM 175
The effects of positive selection on the allele frequency spectrum 176
The effects of balancing selection on the allele frequency spectrum 177
Detecting selection using the allele frequency spectrum: Tajima's D and related tests 179
Interpreting Tajima's D and related tests of the frequency spectrum 181
Detecting selection using the allele frequency spectrum: Fay and Wu's H 185
The power to detect natural selection using tests of the frequency spectrum 186

DETECTING SELECTION USING PATTERNS OF LINKAGE DISEQUILIBRIUM 189

The effects of balancing selection on linkage disequilibrium 189

The effects of completed sweeps on linkage disequilibrium 190

The effects of partial sweeps on linkage disequilibrium 191

Detecting selection using linkage disequilibrium: Haplotype summary statistics 194

Detecting selection using linkage disequilibrium: Measures of pairwise LD 195

Detecting selection using linkage disequilibrium: Haplotype homozygosity 196

Interpreting tests based on haplotype structure and linkage disequilibrium 200

CAVEATS TO TESTS OF LINKED SELECTION 201

CHAPTER 9

DEMOGRAPHIC HISTORY 203

THE DEMOGRAPHIC HISTORY OF SINGLE POPULATIONS 204

Nonequilibrium demographic histories 204

The effect of changing population size on gene genealogies 206

The effect of changing population size on the allele frequency spectrum 208

The effect of changing population size on other measures of variation 210

Types of data and computational methods used in demographic inference 214

Inferring nonequilibrium population histories 217

THE DEMOGRAPHIC HISTORY OF MULTIPLE POPULATIONS 222

The effect of population structure on gene genealogies 223

The effect of population structure on the allele frequency spectrum 225

Inferring demographic histories for multiple populations 228

Multi-population histories as networks 230
The analysis of multiple populations in space 234

POPULATION GENETICS WITHOUT POLYMORPHISM 239
Ancestral parameters inferred from two species 239
Ancestral parameters inferred from gene tree discordance 240
Introgression inferred from gene tree discordance 244

CAVEATS TO INFERENCES OF DEMOGRAPHIC HISTORY 247

CHAPTER 10
POPULATION GENOMICS 249

GENOME-WIDE SCANS FOR SELECTION 250
Motivating questions 250
Selection versus demography 251
General approaches for identifying targets of selection 254

METHODS FOR CARRYING OUT GENOME-WIDE SCANS FOR SELECTION 258
Detecting selection using a sample from a single population 258
Detecting selection using samples from multiple populations 262

CAVEATS ABOUT NON-INDEPENDENCE IN POPULATION GENOMICS 266

FUTURE APPROACHES 268
The application of machine learning to population genetics 268
Ancient DNA and genetic variation through time 270
Populations of whole assembled genomes 271
Deep population sequencing 273

REFERENCES 275

INDEX 315