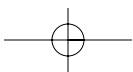
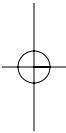
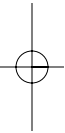
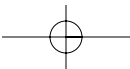
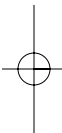
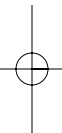


# Elements of Evolutionary Genetics





# Elements of Evolutionary Genetics

BRIAN CHARLESWORTH  
AND  
DEBORAH CHARLESWORTH

ROBERTS AND COMPANY PUBLISHERS  
Greenwood Village, Colorado

**Roberts and Company Publishers**

4950 South Yosemite Street, F2 #197  
Greenwood Village, Colorado 80111 USA  
Internet: [www.roberts-publishers.com](http://www.roberts-publishers.com)  
Telephone: (303) 221-3325  
Facsimile: (303) 221-3326

Publisher: Ben Roberts  
Artist: Emiko-Rose Paul at Echo Medical Media  
Copyeditor: Lee Young  
Production Manager: Betty Gee at Side by Side Studios  
Cover and Interior Designer: Mark Ong at Side by Side Studios

© 2010 by Roberts and Company Publishers

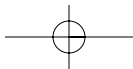
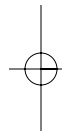
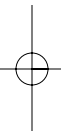
Reproduction or translation of any part of this work beyond that permitted by Section 107 or 108 of the 1976 United States Copyright Act without permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department at Roberts and Company Publishers.

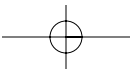
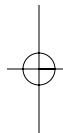
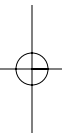
Library of Congress Cataloging-in-Publication Data

Charlesworth, Brian.  
Elements of evolutionary genetics / Brian Charlesworth and Deborah Charlesworth.  
p. cm.  
ISBN 978-0-9815194-2-5  
1. Evolutionary genetics. I. Charlesworth, Deborah. II. Title.  
QH390.C43 2011  
572.8'38—dc22

2009042482

To James F. Crow





# Contents

Preface .....	xxiii
Introduction .....	xxv
<b>1 Variability and its measurement .....</b>	<b>1</b>
Chapter summary .....	1
Introduction .....	2
1.1. Classical and quantitative genetic studies of variability .....	3
1.1.i. Discrete variability .....	3
1.1.ii. Quantitative variability .....	4
1.1.ii.a. Describing quantitative variability .....	4
1.1.ii.b. Genotypes and phenotypes .....	7
1.1.ii.c. The genetic basis of quantitative variability .....	8
1.1.ii.d. Modeling quantitative trait inheritance .....	10
1.1.iii. Concealed variability and mutational variability .....	11
1.1.iii.a. Introduction .....	11
1.1.iii.b. The genetics of inbreeding depression .....	11
1.1.iv. Limitations of the classical and quantitative genetic approaches .....	13
1.2. Molecular variability .....	14
1.2.i. Gel electrophoresis of proteins .....	14
1.2.i.a. Introduction .....	14
1.2.i.b. The discovery of high molecular diversity .....	16
1.2.ii. Detecting variability by restriction mapping .....	20
1.2.iii. DNA sequencing .....	21
1.2.iii.a. Introduction .....	21
1.2.iii.b. DNA sequence polymorphisms; SNPs .....	22
1.2.iii.c. Other types of DNA sequence variants: repeat number polymorphisms and copy number polymorphisms .....	26
1.2.iv. Measuring DNA sequence variability .....	27
1.2.iv.a. Introduction .....	27
1.2.iv.b. Statistics for describing variability .....	27
1.2.v. Haplotypes .....	31
1.2.v.a. Definition and use in measuring variability .....	31
1.2.v.b. The effect of recombination on haplotype variability .....	32
1.3. The causes and maintenance of variability .....	33

1.3.i. Conservation of variability by the mechanism of inheritance	33
1.3.i.a. Hardy–Weinberg equilibrium	34
1.3.i.b. Gene conversion and mutation	35
1.3.ii. Nonrandom mating and identity by descent	35
1.3.ii.a. Identity by descent and coalescence of alleles in common ancestors	36
1.3.ii.b. Genotype frequencies with inbreeding and assortative mating	39
1.3.ii.c. Regular systems of mating	40
1.3.iii. Mutation: the source of new variants	42
1.3.iii.a. The nature of mutations	42
1.3.iii.b. Mutation rates	43
1.3.iii.c. Mutation and evolutionary change	43
Problems	45
<b>2 Basic selection theory and the maintenance of variation</b>	<b>47</b>
Chapter summary	47
Introduction	48
2.1. Basic selection theory: constant fitnesses	49
2.1.i. Haploid populations	49
2.1.ii. Diploid populations	52
2.1.ii.a. Introduction	52
2.1.ii.b. Directional selection	54
2.1.ii.c. Heterozygote advantage and the maintenance of variation	56
2.1.ii.d. Biological implications of heterozygote advantage	59
2.1.ii.e. Mean fitness	61
2.1.ii.f. Heterozygote disadvantage	63
2.2. Variable fitnesses	64
2.2.i. Frequency-dependent selection	65
2.2.i.a. Theory	65
2.2.i.b. Examples of frequency dependence	65
2.2.ii. Fitnesses that vary in time	72
2.2.iii. Spatial variation in fitnesses, hard versus soft selection, and speciation	77
2.2.iv. The biochemical basis of fitness differences	81
Problems	83
<b>3 Directional selection and adaptation</b>	<b>85</b>
Chapter summary	85

## CONTENTS

ix

Introduction	86
3.1. Directional selection on a pair of allelic variants	88
3.1.i. Introduction	88
3.1.ii. Haploid populations	88
3.1.iii. Autosomal inheritance with random mating	90
3.1.iii.a. Introduction	90
3.1.iii.b. Results for weak selection	90
3.1.iii.c. Behavior at extreme frequencies	93
3.1.iii.d. Continuous time	93
3.1.iv. Biological applications	94
3.1.iv.a. Estimating the strength of selection	94
3.1.iv.b. Experimental evolution	95
3.1.iv.c. Haldane's sieve	96
3.1.iv.d. Developmental constraints on evolution	96
3.1.v. More complex selection models: relaxing some of the assumptions	97
3.1.v.a. Autosomal inheritance with sex differences in selection	97
3.1.v.b. Selection on an X-linked gene: Haldane's sieve and sexual antagonism	98
3.1.v.c. Selection in inbreeding populations	100
3.1.v.d. Kin selection	101
3.2. Survival and fixation probability of a favorable mutation	106
3.2.i. The problem	106
3.2.ii. The solution	107
3.2.iii. Biological implications	110
3.3. Selection on quantitative traits	114
3.3.i. Introduction	114
3.3.ii. Quantitative genetic parameters	114
3.3.ii.a. Introduction	114
3.3.ii.b. Components of genetic variance and covariances between relatives	115
3.3.iii. Response to selection	118
3.3.iii.a. Change over one generation	118
3.3.iii.b. Selection coefficients at individual sites	122
3.3.iii.c. Long-term responses to selection	124
3.3.iv. Multiple traits	126
3.4. The evolutionary genetics of adaptation of phenotypes and proteins	129
3.4.i. Introduction	129
3.4.ii. Fisher's geometric model of adaptation	130
3.4.iii. Adaptive evolution of sequences	133
3.4.iii.a. Introduction	133

3.4.iii.b. The distribution of the fitness effects of mutations fixed in an adaptive walk . . . . .	133
3.4.iv. Testing the theory . . . . .	135
3.4.iv.a. The theoretical predictions . . . . .	135
3.4.iv.b. Genetics of interpopulation and interspecies crosses . . . . .	136
3.4.iv.c. Experimental evolution . . . . .	137
3.4.v. Multiple fitness peaks . . . . .	138
Problems . . . . .	141
<b>4 Migration, mutation, and selection . . . . .</b>	<b>143</b>
Chapter summary . . . . .	143
Introduction . . . . .	144
4.1. Migration and selection . . . . .	145
4.1.i. Migration and selection involving two populations . . . . .	145
4.1.ii. Migration and selection in a continuous habitat . . . . .	148
4.1.ii.a. Clines . . . . .	148
4.1.ii.b. Modeling clines . . . . .	150
4.1.ii.c. Using cline widths to measure selection . . . . .	155
4.1.ii.d. Hybrid zones . . . . .	157
4.1.ii.e. The wave of advance of a favorable mutation . . . . .	158
4.2. Mutation and selection . . . . .	159
4.2.i. Introduction . . . . .	159
4.2.ii. Mutation–selection equilibrium . . . . .	160
4.2.ii.a. Autosomes with random mating . . . . .	160
4.2.ii.b. Other cases . . . . .	161
4.2.ii.c. Mutation and selection at multiple sites . . . . .	162
4.2.ii.d. Estimating mutation rates from equilibrium frequencies . . . . .	163
4.3. Genetic load . . . . .	164
4.3.i. Introduction . . . . .	164
4.3.ii. Mutational load . . . . .	164
4.3.ii.a. Autosomal inheritance with random mating . . . . .	164
4.3.ii.b. Other cases . . . . .	165
4.3.ii.c. Multiple sites . . . . .	165
4.3.iii. Segregational load . . . . .	168
4.3.iv. The cost of selection . . . . .	169
4.4. Inbreeding depression . . . . .	170
4.4.i. Introduction . . . . .	170
4.4.ii. Mutation and inbreeding depression in a randomly mating population . . . . .	171
4.4.ii.a. The effects of mutation and selection at a single site . . . . .	171

4.4.ii.b. Multiple sites	171
4.4.ii.c. The inbreeding load	173
4.4.iii. Heterozygote advantage and inbreeding depression	174
4.4.iv. Testing alternative explanations of inbreeding depression	175
4.4.iv.a. Use of $B/A$ ratios	175
4.4.iv.b. Alternative approaches	179
4.4.v. Does dominance evolve?	182
4.5. Genetic variation in fitness components and other quantitative traits	183
4.5.i. Mutational variation	183
4.5.i.a. Direct effects of mutations on fitness	183
4.5.i.b. Mutation and stabilizing selection	186
4.5.ii. Variability maintained by selection	191
Problems	192
<b>5 The evolutionary effects of finite population size:</b>	
<b>basic theory</b>	<b>195</b>
Chapter summary	195
Introduction	196
5.1. Modeling genetic drift	199
5.1.i. Drift without mutation	199
5.1.i.a. The Wright–Fisher model	199
5.1.i.b. Approach to genetic uniformity of a population	202
5.1.i.c. Increase in variance of allele frequencies	204
5.1.ii. Drift and mutation	205
5.1.ii.a. Identity in state of a pair of alleles and the infinite alleles model	205
5.1.ii.b. The infinite sites model	207
5.1.iii. The coalescent process	207
5.1.iii.a. Basic properties	207
5.1.iii.b. The genealogy of two alleles	207
5.1.iii.c. Gene genealogies in general	209
5.1.iii.d. Mutations, the coalescent process, and sequence diversity	209
5.1.iii.e. Simulating the coalescent process	211
5.1.iii.f. Microsatellite loci	215
5.2. Effective population size	216
5.2.i. Introduction	216
5.2.ii. Effective population size and the coalescent process	217
5.2.ii.a. Defining effective population size	217
5.2.ii.b. Calculating $N_e$ with constant population size	218

5.2.iii. Effective population sizes for some biologically interesting cases	.221
5.2.iii.a. Different modes of inheritance (constant population size)	.222
5.2.iii.b. Comparisons of different modes of inheritance	.223
5.2.iii.c. Varying population size	.225
5.2.iii.d. Age- and stage-structured populations and the Moran model	.226
5.2.iii.e. Estimating effective population sizes	.227
5.3. Probability distributions of allele frequencies	.228
5.3.i. Introduction	.228
5.3.ii. The Wright–Fisher model with mutation and selection	.229
5.3.ii.a. The transition from one generation to the next	.229
5.3.ii.b. The probability distribution of variant frequencies	.230
5.3.ii.c. Diffusion equations	.231
5.3.iii. Allele frequency distributions: beyond the Wright–Fisher model	.231
5.3.iii.a. Variance effective population size	.231
5.3.iii.b. Solving the forward diffusion equation	.232
5.3.iii.c. The equilibrium distribution with mutation between neutral variants	.233
5.3.iii.d. The equilibrium distribution with mutation and selection	.239
5.3.iii.e. Biological implications	.240
Appendix to Chapter 5: Diffusion equations	.241
5A.1. General formulation	.241
5A.2. Determining the equilibrium distribution of allele frequencies	.242
Problems	.243
<b>6 Molecular evolution and variation</b>	<b>.245</b>
Chapter summary	.245
Introduction	.246
6.1. Measuring rates of DNA sequence evolution	.247
6.1.i. Comparing a pair of sequences	.247
6.1.ii. Estimating rates of divergence	.248
6.1.ii.a. Basic principles	.248
6.1.ii.b. Correcting for multiple changes	.250
6.1.ii.c. Stochasticity of the substitution process	.251
6.1.ii.d. Correcting for reverse and multiple changes	.252

## CONTENTS

xiii

6.1.ii.e. Nonsynonymous and synonymous changes . . . .	254
6.1.ii.f. Approximate rate constancy of molecular sequence evolution . . . . .	256
6.2. The causes of DNA sequence evolution . . . . .	257
6.2.i. Introduction . . . . .	257
6.2.ii. Neutral sequence evolution . . . . .	257
6.2.ii.a. Expectations for sequence divergence under neutrality . . . . .	257
6.2.ii.b. Fixation of neutral mutations and divergence of sequences from closely related species . . . . .	258
6.2.ii.c. Testing the neutrality of sequence evolution . . . .	260
6.2.iii. Selection in finite populations . . . . .	261
6.2.iii.a. Rates of substitution for unique mutations . . . .	261
6.2.iii.b. Rates of substitution for recurring mutations . . .	264
6.2.iv. Implications for molecular evolution and for studying selection . . . . .	264
6.2.v. Detecting positive selection when purifying selection also acts: codon-based models . . . . .	266
6.3. The distribution of variant frequencies under mutation, selection, and drift . . . . .	268
6.3.i. Introduction . . . . .	268
6.3.ii. Statistical equilibrium under mutation, selection, and drift . . . . .	270
6.3.ii.a. The general model . . . . .	270
6.3.ii.b. The equilibrium frequency distribution under neutrality . . . . .	272
6.3.ii.c. Properties of the statistical equilibrium with selection: rates of evolution . . . . .	275
6.3.ii.d. Properties of the statistical equilibrium with selection: variability . . . . .	277
6.4. Testing for selection from polymorphism and divergence data . .	279
6.4.i. Introduction . . . . .	279
6.4.ii. The McDonald–Kreitman (MK) test . . . . .	280
6.4.ii.a. Introduction . . . . .	280
6.4.ii.b. Principle of the MK test . . . . .	281
6.4.ii.c. Potential difficulties with the MK test . . . . .	284
6.4.ii.d. Estimating the fraction of amino acid fixations caused by positive selection . . . . .	285
6.4.iii. Tests based on variant frequencies . . . . .	287
6.4.iii.a. Introduction . . . . .	287
6.4.iii.b. Tajima's $D$ . . . . .	288
6.4.iii.c. Fu's and Li's test statistics . . . . .	289
6.4.iii.d. Other test statistics . . . . .	290

6.4.iii.e. Problems of interpretation of tests using variant frequencies	290
6.4.iv. Estimating selection intensities	291
6.4.iv.a. Introduction	291
6.4.iv.b. Fitting the distribution of variants to a one-way mutation model	292
6.4.iv.c. Other methods and implications of the results	294
6.4.iv.d. Using polymorphism and divergence statistics	296
Appendix to Chapter 6: More diffusion equation results	297
6A.1. Fixation probabilities	297
6A.2. The equilibrium distribution of variant frequencies with one-way mutations	299
Problems	301
<b>7 Genetic effects of spatial structure</b>	<b>303</b>
Chapter summary	303
Introduction	304
7.1. Describing population differentiation	305
7.1.i. Use of genetic markers to describe differences among populations	305
7.1.i.a. Variances in allele frequencies and $F_{ST}$	305
7.1.i.b. Gene diversities for allelic variants	308
7.1.i.c. Nucleotide site diversity	309
7.1.i.d. Other methods	309
7.1.ii. Empirical patterns	310
7.2. The theory of neutral variation with spatial structure	310
7.2.i. General principles	313
7.2.i.a. Introduction	313
7.2.i.b. Fast time scales and migration models	313
7.2.ii. Mean coalescence times for migration models	315
7.2.ii.a. Introduction	315
7.2.ii.b. The island model	317
7.2.ii.c. Stepping-stone models	320
7.2.ii.d. Mean coalescence times for general migration models	322
7.2.ii.e. The strong migration limit	328
7.2.iii. Continuous populations	329
7.2.iii.a. General considerations	329
7.2.iii.b. Genetic differentiation in continuous populations	332
7.2.iv. Colonization and extinction models	333
7.2.iv.a. Introduction	333
7.2.iv.b. The exchangeable deme model	334

7.2.v. Sampling of alleles from spatially structured populations	338
7.2.v.a. Introduction	338
7.2.v.b. The large deme number approximation	340
7.2.v.c. Scattering and collecting phases	341
7.2.v.d. Within-population coalescence times	347
7.2.vi. The effects of population history	348
7.2.vi.a. Introduction	348
7.2.vi.b. Distinguishing historical changes from stable structure	348
7.3. The theory of selection with spatial structure	350
7.3.i. Introduction	350
7.3.ii. Fixation probabilities in structured populations	350
7.3.iii. Variant frequency distributions with selection and spatial structure	353
7.3.iii.a. The infinite-deme island model	353
7.3.iii.b. Quantitative traits and $Q_{ST}$ values	359
7.3.iii.c. The sampling properties of alleles under the large deme number approximation	360
7.3.iv. Effects of population structure on mean fitness and selection among demes	361
7.3.iv.a. The effects of drift on population mean fitness	361
7.3.iv.b. Inter-deme selection	362
Problems	364
<b>8 Multiple sites and loci</b>	<b>366</b>
Chapter summary	366
Introduction	368
8.1. Describing the state of a population	368
8.1.i. Haplotype frequencies	368
8.1.ii. Measures of linkage disequilibrium	370
8.1.ii.a. Pairs of sites	370
8.1.ii.b. Multiple sites	373
8.1.iii. Empirical results on linkage disequilibrium	374
8.1.iii.a. Observations based on classical genetics	374
8.1.iii.b. Results from electrophoretic surveys	376
8.1.iii.c. Results from surveys of DNA sequence variability	376
8.1.iii.d. Mapping disease genes and quantitative trait loci	378
8.2. Causes of linkage disequilibrium: neutral sites	378
8.2.i. Introduction	378
8.2.ii. Neutral sites at two loci in an infinite population	379

8.2.ii.a. Random mating	379
8.2.ii.b. Inbreeding populations	382
8.2.iii. The effects of finite population size on linkage disequilibrium	383
8.2.iii.a. Linkage disequilibrium and drift in a panmictic population	383
8.2.iii.b. Linkage disequilibrium and the coalescent process	384
8.2.iii.c. Estimating recombination rates	386
8.2.iii.d. The effects of spatial structure	387
8.2.iii.e. Other factors affecting linkage disequilibrium	389
8.3. Neutral and nearly neutral variants linked to selected sites	390
8.3.i. Introduction	390
8.3.ii. The effects of balancing selection	393
8.3.ii.a. The effects of balancing selection on neutral variability	393
8.3.ii.b. Associative overdominance caused by heterozygote advantage	396
8.3.ii.c. Testing for the signature of balancing selection: the HKA test	397
8.3.ii.d. Testing for balancing selection: $D_T$ , $D_F$ , and haplotype tests	398
8.3.iii. Background selection	398
8.3.iii.a. Reduced neutral variability caused by linked deleterious mutations	398
8.3.iii.b. Associative overdominance due to linkage disequilibrium with deleterious mutations	403
8.3.iii.c. Associative overdominance due to identity disequilibrium	404
8.3.iv. The joint effects of spatial and genetic structure	406
8.3.v. Hitchhiking by favorable mutations (selective sweeps)	407
8.3.v.a. Effects on levels of neutral variability at linked sites	407
8.3.v.b. Effects on variant frequency spectra	413
8.3.v.c. Selective sweeps with recombination	415
8.3.v.d. Partial sweeps	416
8.4. Selection at multiple sites in the genome	418
8.4.i. Introduction	418
8.4.ii. Two-site systems	418
8.4.ii.a. General description	418
8.4.ii.b. Two-site fitness matrices	421
8.4.ii.c. Properties of stable equilibria	423
8.4.ii.d. Weak selection and loose linkage	425

8.4.iii. Multisite systems . . . . .	429
8.4.iii.a. Properties of equilibria . . . . .	429
8.4.iii.b. Dynamics under QLE . . . . .	430
8.4.iv. The evolutionary dynamics of quantitative traits . . . . .	430
8.4.iv.a. The infinitesimal model . . . . .	430
8.4.iv.b. Extending the infinitesimal model . . . . .	434
8.4.v. Wright's shifting balance model of evolution . . . . .	434
8.4.v.a. Epistasis and multiple peaks of mean fitness . . . . .	434
8.4.v.b. The fixation of chromosomal rearrangements . . . . .	436
8.4.vi. The evolution of hybrid fitness loss . . . . .	437
Appendix to Chapter 8: Linkage disequilibrium at neutral sites in finite populations . . . . .	439
8A.1. The general method . . . . .	439
8A.2. Application to linkage disequilibrium . . . . .	440
Problems . . . . .	442

<b>9 The evolution of breeding systems, sex ratios, and life histories . . . . .</b>	<b>444</b>
Chapter summary . . . . .	444
Introduction . . . . .	445
9.1. Sexual versus asexual reproduction; inbreeding versus outcrossing . . . . .	447
9.1.i. Introduction . . . . .	447
9.1.ii. Asexual versus sexual reproduction . . . . .	447
9.1.ii.a. Introduction . . . . .	447
9.1.ii.b. The cost of sex . . . . .	447
9.1.ii.c. The maintenance of sexual reproduction . . . . .	449
9.1.iii. Sexual systems and mating systems . . . . .	450
9.1.iii.a. Sexual systems . . . . .	450
9.1.iii.b. Mating systems . . . . .	450
9.1.iii.c. Self-incompatibility . . . . .	452
9.1.iii.d. Detecting outcrossing and estimating outcrossing rates in natural populations . . . . .	453
9.2. The evolution of inbreeding and outbreeding . . . . .	455
9.2.i. Introduction . . . . .	455
9.2.ii. The evolution of selfing rates . . . . .	456
9.2.ii.a. Introduction . . . . .	456
9.2.ii.b. The transmission advantage of selfing (the cost of outcrossing) . . . . .	458
9.2.ii.c. Why do many organisms outcross? Inbreeding depression and the evolution of selfing rates . . . . .	458
9.2.ii.d. More realistic models of inbreeding depression; purging of deleterious mutations . . . . .	464

9.2.ii.e. Origins and evolution of self-incompatibility systems	465
9.2.iii. The breakdown of self-incompatibility systems	466
9.2.iii.a. The breakdown of heterostyly	466
9.2.iii.b. The breakdown of homomorphic SI	467
9.3. The evolution of separate sexes	467
9.3.i. Introduction	467
9.3.ii. The evolutionary stability of cosexuality	468
9.3.iii. Pathways to dioecy from cosexuality	468
9.3.iv. When will unisexuals invade cosexual populations?	471
9.3.iv.a. The establishment of cytoplasmic male sterility factors in populations	471
9.3.iv.b. The establishment of nuclear male sterility factors	472
9.3.iv.c. Testing the theory	474
9.3.iv.d. Androdioecy	476
9.3.v. The transition from gynodioecy to dioecy	478
9.3.v.a. The effect of the presence of females	478
9.3.v.b. The importance of linkage	480
9.4. Evolutionarily stable strategies, sex ratios, and sex allocation	483
9.4.i. Introduction	483
9.4.ii. Evolutionarily Stable Strategies	484
9.4.iii. The ESS sex ratio in a panmictic population	486
9.4.iii.a. Equal costs of male and female offspring	486
9.4.iii.b. Unequal costs of sons and daughters	486
9.4.iv. Extraordinary sex ratios	488
9.4.iv.a. Local mate competition	488
9.4.iv.b. Haplodiploid eusocial insects	490
9.4.v. Sex allocation in cosexuals	492
9.4.v.a. Theoretical models	492
9.4.v.b. Empirical tests	493
9.5. Age-structured populations and the evolution of life histories	494
9.5.i. Introduction	494
9.5.ii. Demographic models	494
9.5.ii.a. Basic concepts and notation	494
9.5.ii.b. Predicting population growth	496
9.5.ii.c. Population genetics for age-structured populations	502
9.5.iii. Selection equations	502
9.5.iii.a. The general recursion relations with no sex differences	502

## CONTENTS

xix

9.5.iii.b. Equilibrium conditions . . . . .	504
9.5.iii.c. Stability and invasion conditions . . . . .	505
9.5.iii.d. Weak selection approximations . . . . .	506
9.5.iii.e. Relationships between different fitness measures . . . . .	507
9.5.iii.f. More general situations . . . . .	508
9.5.iv. Implications for population genetic studies . . . . .	508
9.5.v. Principles of life history evolution . . . . .	510
9.5.v.a. The sensitivity of fitness to changes in demographic parameters . . . . .	510
9.5.v.b. Trade-off models . . . . .	512
9.5.v.c. Other life history models . . . . .	514
9.5.v.d. The evolution of aging . . . . .	515
Problems . . . . .	519
<b>10 Some topics in genome evolution . . . . .</b>	<b>521</b>
Chapter summary . . . . .	521
Introduction . . . . .	522
10.1. Evolution of noncoding sequences and codon usage . . . . .	523
10.1.i. Introduction . . . . .	523
10.1.i.a. Base composition . . . . .	523
10.1.i.b. Codon usage bias . . . . .	524
10.1.ii. Evolutionary forces affecting base composition . . . . .	525
10.1.ii.a. Introduction . . . . .	525
10.1.ii.b. Mutation and genetic drift . . . . .	526
10.1.ii.c. Lower effectiveness of selection with reduced recombination . . . . .	527
10.1.ii.d. Biased gene conversion . . . . .	527
10.1.ii.e. Departures from equilibrium base composition . . . . .	529
10.1.iii. Evolutionary forces affecting codon usage . . . . .	530
10.1.iii.a. Measuring codon usage bias . . . . .	530
10.1.iii.b. Causes of selection for codon usage bias . . . . .	533
10.1.iv. Estimating the intensity of selection or biased gene conversion . . . . .	535
10.1.iv.a. The intensity of selection and BGC on noncoding sites . . . . .	536
10.1.iv.b. The intensity of selection on codon usage . . . . .	537
10.2. The evolutionary causes and consequences of genetic recombination . . . . .	539
10.2.i. Introduction . . . . .	539
10.2.ii. The cost of recombination . . . . .	540

10.2.iii. The evolution of close linkage	543
10.2.iii.a. Inversion polymorphisms	543
10.2.iii.b. Supergenes	543
10.2.iii.c. Centromeric and telomeric suppression of crossing over	544
10.2.iii.d. Congealing of the genome versus selection for recombination	545
10.2.iv. Possible advantages of sex and recombination: deterministic mechanisms	546
10.2.iv.a. Introduction	546
10.2.iv.b. Fluctuating epistasis	548
10.2.iv.c. Moving optima	548
10.2.iv.d. Mutation and selection	549
10.2.v. Stochastic processes	550
10.2.v.a. General considerations	550
10.2.v.b. Selective sweeps	551
10.2.v.c. Background selection	552
10.2.v.d. Muller's ratchet	552
10.2.v.e. Reverse mutation and Hill–Robertson interference	556
10.2.vi. Empirical evidence concerning the consequences of recombination	556
10.2.vi.a. Reduced levels of adaptation in regions with low recombination	556
10.2.vi.b. Discriminating between the different possible advantages of recombination	559
10.3. Selfish DNA	561
10.3.i. Introduction	561
10.3.ii. Gametic selection and segregation distortion	561
10.3.ii.a. Gametic selection	561
10.3.iii. Meiotic drive and segregation distortion	562
10.3.iii.a. Introduction	562
10.3.iii.b. Segregation distortion	563
10.3.iii.c. Examples of segregation distorter systems	565
10.3.iii.d. The evolution of segregation distorter systems	568
10.3.iv. Transposable elements	571
10.3.iv.a. Introduction	571
10.3.iv.b. The evolution of TEs in genomes	573
10.3.iv.c. TE behavior in genomes and populations	574
10.3.iv.d. Population dynamics of TEs	576

CONTENTS

10.3.iv.e. Discriminating among possible regulators of element abundance	580
10.3.iv.f. Genomes and genomic regions with high TE contents	584
Problems	585
<b>Appendix</b>	<b>587</b>
A1. Basic mathematics	587
A1.i. Differentiation	587
A1.ii. Representing functions by series	588
A1.ii.a. Taylor's series	588
A1.ii.b. Approximate numerical solutions of equations	589
A1.ii.c. Some useful series	589
A1.iii. Integration	590
A1.iv. Vectors and matrices	591
A1.iv.a. Some definitions	591
A1.iv.b. The inverse of a matrix and the solution of linear equations	593
A1.iv.c. Eigenvalues and eigenvectors	594
A1.iv.d. The spectral expansion of a matrix	596
A2. Basic probability and statistics	597
A2.i. Introduction	597
A2.ii. The rules of probability	597
A2.iii. The moments of a distribution	598
A2.iv. Pairs of variables	599
A2.v. Some important probability distributions	601
A2.v.a. The binomial distribution	601
A2.v.b. The multinomial distribution	602
A2.v.c. The geometric and exponential distributions	602
A2.v.d. The Poisson distribution	603
A2.v.e. The normal distribution	603
A2.v.f. Other distributions	605
A2.vi. Statistical estimation	605
A2.vi.a. Introduction	605
A2.vi.b. Maximum likelihood	606
A2.vi.c. Bayesian methods	608
<b>Answers to problems</b>	<b>609</b>
Chapter 1	609
Chapter 2	611
Chapter 3	614
Chapter 4	618

xxii CONTENTS

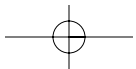
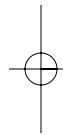
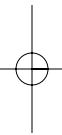
Chapter 5 .....	621
Chapter 6 .....	624
Chapter 7 .....	629
Chapter 8 .....	633
Chapter 9 .....	637
Chapter 10 .....	640
References .....	642
Author Index .....	715
Subject Index .....	723

# Preface

The project of writing a textbook on evolutionary genetics was suggested to us by Ben Roberts over three years ago. We are very grateful to him for his support and encouragement over the seemingly endless process of writing and revision, and to Emiko Paul for turning our crude sketches and blurred images into handsome illustrations. We also thank Jane Charlesworth for assistance with the index, and Betty Gee and Mark Ong for overseeing the production of the book.

We especially thank the people who generously gave their time to read and comment on some or all of the draft chapters: Hiroshi Akashi, Jane Charlesworth, Jim Crow, George Gilchrist, Philip Johnson, Laurence Loewe, Christina Muirhead, Sally Otto and her lab group, Manus Patten, Jitka Polechova, Suo Qiu, Sohini Ramachandran, Monty Slatkin, and Kai Zeng. They corrected numerous errors and made many suggestions for improving the contents of the book. We have not always taken their advice, and the responsibility for the remaining errors and lapses from clarity is entirely ours.

*Edinburgh, November 2009*



# Introduction

“Evolution” means cumulative change over time in the characteristics of a population of living organisms. Evolutionary changes have been in progress at all times, including the present. Some can be very fast. A pest or bacterial population susceptible to insecticide or antibiotic can be transformed into a resistant population in a few generations, under the strong selection pressures that operate when a pesticide or drug is introduced into use. Many evolutionary changes, however, are very slow, and take place over millions of years; for example, the emergence of mammals from reptiles. Events such as the evolution of a new species lie in the middle of this time spectrum. The types of traits involved in evolutionary change are enormously varied, ranging from easily observable characteristics of organisms, such as visible color patterns or behavior, to the components of the genome itself. Understanding evolutionary change in the sequences of the bases that make up the genetic information carried by DNA is a major focus of modern biology.

All evolutionary changes require initially rare genetic variants to spread among the members of the population, rising to a high frequency, so that the population becomes genetically different from its ancestral condition. The goal of evolutionary geneticists is to understand such changes, and to explain what has happened in evolution. We now largely understand the origin of the variants that are the basis of evolutionary changes, and the mechanisms that can cause variants to spread through populations. As in other branches of science, we assume that the properties of organisms living in the past were fundamentally the same as for present-day living organisms, so that even remote evolutionary events were subject to the rules revealed by experimental studies today. This assumption is supported by the basic similarities of the genetic material and the genetic code across all types of organisms alive today, even ones whose last common ancestor lived more than two billion years ago.

Evolutionary genetics differs from most other branches of biology in its strong theoretical structure. The rules of transmission of genetic information from parents to offspring, uncovered by experimental genetics, impose important constraints on what can happen in evolution. From the very beginnings of genetics, early in the 20th century, mathematical models of genetic processes in populations have been used to study evolution. A major goal of this book is to show how these models help us understand evolution, and how they can add insights beyond those that we can gain intuitively, but sometimes giving counterintuitive results. In addition, they can provide results that are far from obvious, and provide a quantitative understanding that leads to new ways of asking and answering important biological questions.

In this book, therefore, we present the models in the context of biological questions, and analyze them only to a level of detail needed to illuminate these questions. Our aim is keep our book as short as possible, and to provide a foundation of knowledge of the main concepts of evolutionary genetics for use by people with a variety of biological interests. We assume familiarity with the basic principles of genetics and molecular biology; readers with backgrounds in fields other than biology will first need to learn these principles from an introductory text.

In each chapter, the simpler and more basic concepts are described first, and advanced topics are usually presented either at the ends of some chapters (where, as we indicate, they can be skipped, because most are not essential for understanding later chapters), or in **Appendices (Chapters 5, 6, and 8)**. Some special topics illustrating important principles are discussed as separate **Case Studies**. Each major section of the book (indicated by arabic numerals) could have been a separate chapter, but we have grouped related topics into chapters, with divisions into major sections providing natural breaks that we hope will help readers to pause and think, and make sure that they have understood the ideas they have been reading about. Slow reading and working through many sections will be necessary, because they contain concepts and approaches that will be new to most readers.

Our style of derivation is not rigorous. As J. B. S. Haldane (1964) wrote in his famous essay *A defense of beanbag genetics*:

... Wright, Fisher, and I all made simplifying assumptions which allowed us to pose problems soluble by the elementary mathematics at our disposal. ... Our mathematics may impress zoologists, but do not greatly impress mathematicians.

To avoid disrupting the flow of the verbal arguments, the mathematical derivations are separated into boxes, while the text discusses and illustrates the main conclusions that can be drawn from them. To help readers understand the approaches and concepts, problems are also provided, with worked answers at the end of the book. The problems range from simple numerical examples (intended to help readers see how a concept is used), to short derivations (to help readers understand how the derivations work). Working through the boxes and the problems will give a good understanding of the basics of the theory, but the text can be read without reading the boxes (e.g., when first reading a section of the book). Most of the boxes and problems are accessible to any reader who can do simple algebra, but some (indicated by stars) require basic calculus or matrix algebra. A summary of some useful basic results in mathematics and statistics is given in the final **Appendix**.

This book is appropriate for graduate students in evolutionary biology and human genetics (or students in advanced level undergraduate courses), and for

researchers in related fields who wish to become acquainted with evolutionary genetics. Because of the emphasis on specific biological questions, individual sections of the book can also be used in general evolution or genetics courses.

We illustrate the questions illuminated by the theory with examples drawn from a wide variety of organisms, ranging from viruses to humans. In the earlier, more basic, chapters we emphasize the historical development of the field. Classical population genetics is classic science, with beautiful and often simple results, providing the basis for an understanding of evolution and its genetic basis that should be part of every biologist's consciousness. **Chapter 1** describes the evidence accumulated from over a century of work on the nature and amount of genetic variability in populations. Recent advances in DNA sequencing are revolutionizing the collection of data on differences between species and among individuals within species. This is providing evolutionary genetics with an unprecedented wealth of material, to which the models developed by the pioneers of population genetics are often surprisingly relevant. Furthermore, the great power of genome-wide surveys of populations to detect associations between genetic variants and diseases or complex traits, and to test for the effects of selection, means that evolutionary genetics is more relevant than ever before to practical problems, such as animal and plant improvement and human health.

The major processes that cause variability within species and populations to be incorporated into evolutionary change are *natural selection* (differences in reproductive success among individuals with different characteristics) and *genetic drift* (the random sampling of variants during the transmission of genes from one generation to the next). These processes, and their interactions with other evolutionary forces such as *mutation*, *migration*, and *recombination*, are described in **Chapters 2–8**. The final two chapters (**9** and **10**) use these concepts to study some evolutionary questions for which genetic approaches are central.

The book does not comprehensively cover the whole of evolutionary genetics, which has now developed into a major research area in biology, some of it highly mathematical. Rather, as explained above, our aim is to describe basic population genetic processes and to illustrate their relevance to interesting evolutionary situations. For this reason, we only touch on the rapidly growing field of statistical inference in population genetics. Many other interesting results in population genetics and evolution are also omitted, as are details of applications to medical genetics and plant and animal improvement. We hope that our choice of topics represents the major concepts in the field, but such judgments are subjective, and reflect our own work and interests. Where possible, we provide references to further work, either work on topics related to those we discuss, or more mathematical developments of the same topics.

