PROBLEMS AND ANSWERS

CHAPTER 2

Problems

1. A population of infants from Musoma, Tanzania, was scored for the β-hemoglobin locus genotype (determined by a single autosomal locus with two alleles, *A*and *S*) as follows:

Genotype	*AA*	*AS*	*SS*	Total
Number	189	89	9	287

Characterize this population by its genotypic frequencies. Characterize the gene pool by the allele frequencies for *A* and *S*. Using the Hardy–Weinberg law, predict the genotypic frequencies.

2. Among people of southern Italian and Sicilian ancestry living in Rochester, New York ($N = 10{,}000$), about one birth in 2500 has thalassemia major (a type of anemia) and about one birth in 25 has a milder anemia known as thalassemia minor. Are these data compatible with a single-locus hypothesis as a basis for the heredity of these anemic conditions? Why or why not?

3. Calculate the genotype frequencies expected under Hardy-Weinberg to determine which of the following populations are at Hardy-Weinberg genotypic frequencies. For any population not at Hardy-Weinberg, state whether there is an excess or a deficiency of heterozygotes.

Genotype	*AA*	*Aa*	*aa*	Total
a.	50	20	30	100
b.	25	10	1	36
c.	20	20	5	45
d.	9	10	81	100
e.	5625	3750	625	10,000

4. Two populations are examined for the same gene locus with the following results:

Population 1: *AA* 162; *Aa* 36; *aa* 2

Population 2: *AA* 18; *Aa* 84; *aa* 98

Population Genetics and Microevolutionary Theory, By Alan R. Templeton

Now suppose these two populations are combined to form a new population:

Population 3: *AA* 180; *Aa* 120; *aa* 100

a. What system(s) of mating do populations 1 and 2 have for this locus? Are their different genotype frequencies explained by different systems of mating? If not, why do their genotype frequencies differ?

b. Is population 3 at Hardy–Weinberg equilibrium? If not, indicate whether heterozygotes are in excess or in deficiency.

c. For how many generations will the effect of this single episode of admixture of populations be detectable in a population established from population 3 with respect to genotype frequencies if mating is at random?

5. A population is polymorphic for an *Eco*R1 restriction site in a defined chromosomal region. Of the chromosomes 75% have the cut site and 25% do not. A closely linked mutation then occurs that causes a *Hin*dIII cut site to appear. This mutation occurs on a chromosome without the *Eco*R1 cut site. With subsequent evolution, the frequency of the *Hin*dIII cut site evolves to a current value of 0.1, although the frequency of the *Eco*R1 cut site remains at 0.75. During this time, no recombination events occur between the *Eco*R1 and *Hin*dIII sites.

a. Calculate the gametic-phase imbalance (linkage disequilibrium) between the two restriction sites in the current population.

b. Suppose that the initial and current frequencies of the *Eco*R1 cut site had both been 0.2 instead of 0.75 but all else is as described above. Calculate the gametic-phase imbalance between the two restriction sites in the current population.

6. One thousand individuals were typed for two traits determined by two different autosomal loci, each with two alleles (A and a; B and b) with the following table giving the two-locus genotype numbers:

	BB	*Bb*	*Bb*	Sum
AA	36	48	16	100
Aa	72	96	32	200
Aa	252	336	112	700
Sum	360	480	160	1000

a. Determine if *each locus* is at a single-locus Hardy–Weinberg equilibrium. If not, what would the genotype frequencies be at each locus considered separately?

b. Are these two loci in linkage equilibrium? If not, what would each gametic combination frequency be if they were in linkage equilibrium?

c. Is this array of nine genotypes consistent with random mating and neutrality? If not, what would be the expected frequencies of these nine biloci genotypes?

7. The frequency of phenylketonuria (PKU, caused by an autosomal recessive allele) is 0.00004 at birth. Assuming Hardy–Weinberg, what is the frequency of the PKU allele?

What is the expected Hardy–Weinberg ratio of PKU carriers (heterozygotes) to affecteds (PKU homozygotes)?

8. Given random mating and the absence of mutation migration and selection, what will be the genotype frequencies in the zygotes of the next generation of a very large population if the initial population had the following genotype frequencies?

Genotype	*AA*	*Aa*	*aa*
a.	0.16	0.48	0.36
b.	0.30	0.00	0.70
c.	0.20	0.40	0.40

9. A person has type O blood only if they are homozygous for the *O* allele at the *ABO* locus. The frequency of the *O* allele is 0.67 in the United States. A person's blood is Rh+ if he or she is homozygous or heterozygous for the *D* allele at the *Rh* locus, which has an allele frequency of 0.6. Assuming no linkage disequilibrium between these two unlinked loci, what is the frequency of people with blood type O+, assuming a random-mating population?

10. Two populations are examined for the same pair of loci with the following results in terms of two-locus gamete numbers:

Population 1: *AB* 720; *Ab* 180; *aB* 80; *ab* 20
Population 2: *AB* 300; *Ab* 2700; *aB* 700; *ab* 6300

These two populations are now combined to form a third:

Population 3: *AB* 1020; *Ab* 2880; *aB* 780; *ab* 6320

a. Calculate the two-locus gamete frequencies for each population and the gametic-phase imbalance (linkage disequilibrium). Is any population out of linkage equilibrium?

b. What is the effect of pooling of the two populations on linkage equilibrium?

c. Suppose that the recombination rate between loci *A/a* and *B/b* is 0.1. What is the expected disequibrium in the offspring of population 3 (the parental gene pool), assuming random mating? In the second generation?

d. Can this single episode of admixture be detected in the population established from population 3 after two generations of random mating? Can it be detected in the genotype frequencies at the *A/a* locus after two generations of random mating? Can it be detected in the genotype frequencies at the *B/b* locus after two generations of random mating?

11. Consider two unlinked loci each with two alleles: *A* and *a*; *B* and *b*. The frequencies of the nine observable genotypes in a population with respect to these loci are:

	AA	*Aa*	*aa*
BB	0.20	0.40	0.04
Bb	0.05	0.07	0.20
bb	0.00	0.03	0.01

- **a.** Is this population at Hardy–Weinberg equilibrium for each locus considered separately?
- **b.** Is this population at Hardy–Weinberg equilibrium with no gametic-phase imbalance with respect to the nine, two-locus genotypic frequencies?
- **c.** If it is not in two-locus equilibrium, what are the equilibrium genotype frequencies and how long will it take to dissipate the gametic-phase imbalance to less than 5% of its initial value?
- **d.** If the loci are linked with a recombination frequency of 0.001, how will this change the equilibrium genotype frequencies?

12. A population of Greenland Eskimos, when tested for the blood antigens M and N (determined by a single locus with two codominant alleles), was categorized as follows:

Blood type	M	MN	N	Total
Number	61	64	27	152

- **a.** Characterize this population by its genotypic frequencies.
- **b.** Characterize the gene pool by the allele frequencies for M and N.
- **c.** Using the Hardy–Weinberg law, predict the genotypic frequencies.
- **d.** Test the goodness of fit of this population to the Hardy–Weinberg expectations.

13. Suppose that in the above population it was not possible to test for the M antigen, but instead it was only possible to test for the presence or absence of antigen N.

- **a.** Assuming Hardy–Weinberg, what is the frequency of the allele coding for the M antigen?
- **b.** What is the expected Hardy–Weinberg ratio of *MN* heterozygotes to *MM* homozygotes?

14. A person has type M blood only if he or she is homozygous for the *M* allele at the *MN* locus. The frequency of the *M* allele is 0.24 in European Americans and 0.09 in African Americans. A person's blood is Rh+ if homozygous or heterozygous for the *D* allele at the *Rh* locus, which has an allele frequency of 0.6 in European Americans and 0.84 in African Americans.

- **a.** Assuming no linkage disequilibrium between these two unlinked loci within European Americans and within African Americans, what are the frequencies of gametes bearing the *M* and *D* alleles in European Americans and African Americans?
- **b.** Suppose all the people in a town are surveyed for these alleles, and the population consists of 70% European Americans and 30% African Americas. Do the *M*and *D* alleles show linkage disequilibrium in the total town population, and if so, what is its value?
- **c.** Assuming random mating within European Americans and African Americans, what is the frequency of Rh+ people in the total town population? What is the frequency of the *D* allele in the total town population? Is the frequency of Rh+ people in the total town population predicted well by applying the Hardy–Weinberg law to the total town gene pool (don't test statistically, just look at exact numbers)?
- **d.** Suppose the town population becomes a single random-mating population. What is the frequency of Rh+ people in the town after a single generation of random mating? What is the linkage disequilibrium between the *M* and *D* alleles after a

single generation of random mating given that the *MN* and *Rh* loci are on separate autosomes?

15. The disease erythroblastosis fetalis occurs when an Rh− woman has an Rh+ child. If there is some leakage across the placenta, the Rh− mother can produce antibodies against the Rh+ fetus. This usually does not affect the first Rh+ child, but subsequent pregnancies with Rh+ fetuses can be severely affected by this immune reaction. Given the data in problem 14, what are the frequencies of families at risk for this disease in European Americans and in African Americans, assuming random mating within these two populations?

16. Consider a population with the following gamete frequencies at a pair of loci: *AB* 0.4; *Ab* 0.3; *aB* 0.2; *ab* 0.1.
    - **a.** Calculate the linkage disequilibrium as both D and D' (equation 2.15).
    - **b.** Suppose that the recombination rate between loci A/a and B/b is 0.2. What is the expected disequibrium (D) in the next generation, assuming random mating? In the second generation?
    - **c.** Now suppose that the recombination rate is 0.01. Repeat the calculations in part b under this assumption.

Answers

1. First, you must calculate the genotypic frequencies, since your calculations from this point on rely on frequencies, not absolute numbers:

Genotype	*AA*	*AS*	*SS*	Total
Number	189	89	9	287
Frequency	189/287 = 0.66	0.31	0.03	1.0

Next, calculate allele frequencies:

$$p(A) = G_{AA} + \frac{1}{2}G_{AS} = 0.66 + 0.155 = 0.815$$

$$q(S) = G_{SS} + \frac{1}{2}G_{AS} = 0.03 + 0.155 = 0.184$$

Using the values you obtain for p and q above, determine the Hardy–Weinberg genotype frequencies:

AA	*AS*	*SS*
p^2	$2pq$	p^2
0.662	0.303	0.035

2. To answer this question, construct a model with all the elements of the basic Hardy–Weinberg model and see if the data you have been given fit. Assume two alleles and one locus and arbitrarily name the alleles A and a. You have the genotypic frequencies of two genotypes from the question: $Aa = 1/25 = 0.04$; $AA = 1/2500 = 0.0004$.

Therefore:

AA (thalassemia major)	*Aa* (thalassemia minor)	*aa* (no anemia)
0.0004	0.04	0.9596 $(1-G_{AA} - G_{Aa})$

Now, calculate p and q:

$$p = 0.0004 + 0.02 = 0.0204$$

$$q = 0.9596 + 0.02 = 0.9796$$

Given these values for p and q, the expected genotype frequencies under Hardy–Weinberg are $p^2 = 0.0004$; $2pq = 0.04$; $q^2 = 0.9596$. These match the given values in the population. Therefore, this trait fits the Hardy–Weinberg two-allele, one-locus model.

3. These calculations should be done as shown in problem 1:

Genotype	*AA*	*Aa*	*Aa*	Heterozygote
a. Observed number	50	20	30	
Observed frequency	0.5	0.2	0.3	
Hardy–Weinberg expected	0.36	0.48	0.16	Deficiency
b. Observed number	25	10	1	
Observed frequency	0.694	0.278	0.028	
Hardy–Weinberg expected	0.694	0.278	0.028	In Hardy–Weinberg
c. Observed number	20	20	5	
Observed frequency	0.444	0.444	0.111	
Hardy–Weinberg expected	0.444	0.444	0.111	In Hardy–Weinberg
d. Observed number	9	10	81	
Observed frequency	0.09	0.10	0.81	
Hardy–Weinberg expected	0.02	0.24	0.78	Deficiency
e. Observed number	5625	3750	625	
Observed frequency	0.563	0.375	0.063	
Hardy–Weinberg expected	0.563	0.375	0.063	In Hardy–Weinberg

4. a. To determine the system of mating that these populations are experiencing, first calculate their Hardy–Weinberg parameters:

Population 1:	*AA*	*Aa*	*aa*	
	162	36	2	
G_{ij} observed	$162/200 = 0.81$	0.18	0.01	$p = 0.81 + 0.09$
G_{ij} expected	$p^2 = 0.81$	$2pq = 0.18$	$q^2 = 0.01$	$= 0.9; q = 0.1$

Population 2:	AA	Aa	aa	
	18	84	98	
G_{ij} observed	$18/200 = 0.09$	0.42	0.49	$p = 0.09 + 0.21$
G_{ij} expected	$p^2 = 0.09$	$2pq = 0.42$	$q^2 = 0.49$	$= 0.3; q = 0.7$

Both of these populations meet Hardy–Weinberg expectations: Random mating cannot be rejected in either of these populations. The genotype frequencies between populations 1 and 2 differ because of different allele frequencies, not because of different systems of mating.

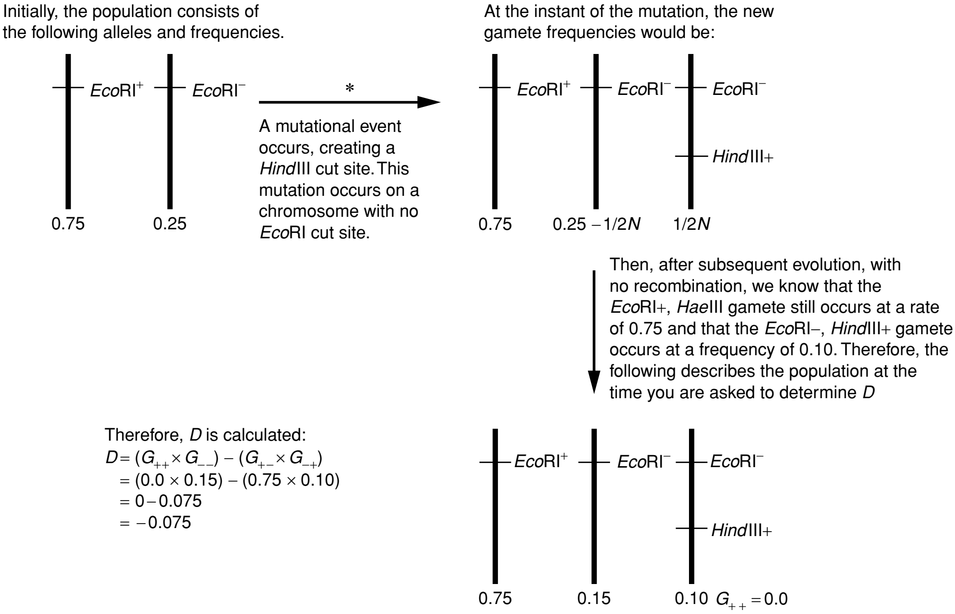
b. Again, first calculate the Hardy–Weinberg parameters:

Population 3:	AA	Aa	aa	
	180	120	100	
G_{ij} observed	$180/400 = 0.45$	0.30	0.25	$p = 0.45 + 0.15$
G_{ij} expected	$p^2 = 0.36$	$2pq = 0.48$	$q^2 = 0.16$	$= 0.6; q = 0.4$

This population does not meet Hardy–Weinberg expectations. The observed frequency of heterozygotes is 0.30, but the Hardy–Weinberg expected frequency is 0.48. Therefore, there is a deficiency of heterozygotes.

c. With random mating after the admixture event, the admixture will be undetectable after the first generation.

5. a. It may be helpful to draw a picture in order to conceptualize this problem correctly:



Initially, the population consists of the following alleles and frequencies.

A mutational event occurs, creating a *Hind*III cut site. This mutation occurs on a chromosome with no *Eco*RI cut site.

At the instant of the mutation, the new gamete frequencies would be:

Then, after subsequent evolution, with no recombination, we know that the *Eco*RI+, *Hae*III gamete still occurs at a rate of 0.75 and that the *Eco*RI−, *Hind*III+ gamete occurs at a frequency of 0.10. Therefore, the following describes the population at the time you are asked to determine D

Therefore, D is calculated:

$D = (G_{++} \times G_{--}) - (G_{+-} \times G_{-+})$

$= (0.0 \times 0.15) - (0.75 \times 0.10)$

$= 0 - 0.075$

$= -0.075$

b. This part is done exactly as part a except that the initial frequencies of the alleles in the population are changed. The final result:

The gamete frequencies are now:

	*Eco*R1$^+$	*Eco*R1$^-$
*Hin*dIII$^+$	0.00	0.10
*Hin*dIII$^-$	0.20	0.70

So, $D = (0.00)(0.70) - (0.10)(0.20) = -0.020$.

6. **a.** Approach this part the same way you did problem 1. For the A locus (row sums), $p = 0.2$ and the Hardy–Weinberg (HW) frequencies are 0.04, 0.32, and 0.64, respectively. For the B locus (column sums), $p = 0.6$ and it is exactly at HW.

b. b. At linkage equilibrium the two-locus gamete frequencies are given by the products of the marginal allele frequencies. Therefore, to test if there really is no linkage disequilibrium, we must determine if the gamete frequencies are equal to the products of the marginal frequencies. We can test this by determining the actual gamete frequencies from the data by summing over gamete types:

	B	b	Total
A	$36 + 1/248 + 1/272 + 1/496 = 120$	80	200
a	$252 + 1/272 + 1/2336 + 1/496 = 480$	320	800
Total	600	400	1000

At linkage equilibrium, the two-locus gamete frequencies are given by the products of the allele (marginal) frequencies, which they are. Therefore, this system has no linkage disequilibrium.

c. Take the product of the marginal HW frequency to get the total HW expected frequencies:

A Locus Genotype	Two-Locus Expected Genotype Frequency			A Locus HW
AA	0.0144	0.0192	0.0064	0.04
Aa	0.1152	0.1536	0.0512	0.32
Aa	0.2304	0.3072	0.1024	0.64
B locus genotype	BB	Bb	Bb	
B locus HW	0.36	0.48	0.16	

These do NOT correspond to the observed frequencies because locus A is not in HW.

7. $q^2 = 0.00004$, so $q = 0.0063$, so $2pq = 0.0126$. Ratio $= 2pq/q^2 = 315 : 1$.

8.

		AA	Aa	aa
a. $p = (.16) + \frac{1}{2}(0.48) = 0.4$	Next generation frequencies:	$(0.4)^2 = 0.16$	$2(0.4)(0.6) = 0.48$	$(0.6)^2 = 0.36$
b. $p = (.30) + \frac{1}{2}(0.00) = 0.3$	Next generation frequencies:	$(0.3)^2 = 0.09$	$2(0.3)(0.7) = 0.42$	$(0.7)^2 = 0.49$
c. $p = (0.20) + \frac{1}{2}(0.40) = 0.4$	Next generation frequencies:	$(0.4)^2 = 0.16$	$2(0.4)(0.6) = 0.48$	$(0.6)^2 = 0.36$

9. The genotype frequency under random mating of OO is $(0.67)^2 = 0.45$. The frequency of Rh+ individuals is $(0.6)^2 + 2(0.6)(0.4) = 0.84$. With no linkage disequilibrium, the frequency of O+ individuals is $(0.45)(0.84) = 0.38$.

10. **a.** $D(1) = (0.72)(0.02) - (0.18)(0.08) = 0$
$D(2) = 0$
$D(3) = (0.09273)(0.57455) - (26{,}182)(0.07091) = 0.03471$
Population 3 is not in linkage equilibrium.

b. Creates linkage disequilibrium.

c. $D_1 = (1 - r)D_0 = 0.9(0.03471) = 0.03124$; $D_2 = 0.9(0.03124) = 0.02812$

d. $D_t \neq 0$ at $t = 2$, so admixture is still detectable as a nonequilibrium state. Admixture is not detectable with the single-locus genotype frequencies (see problem 4).

11. Let p be the frequency of A, q of a, r of B, and s of b.

a. First determine the genotype frequencies for just the A/a locus by adding up the frequencies in the columns and the B/b locus genotype frequencies by adding up the rows. Then determine allele frequencies by adding up appropriate genotype frequencies (dividing by 2 for heterozygotes) to obtain $p = 0.5$, $q = 0.5$, $r = 0.8$, and $s = 0.2$. Both single loci are in Hardy–Weinberg.

b. No. This is immediately apparent from the complete absence of the $AAbb$ genotype.

c. Multiply the appropriate single-locus Hardy–Weinberg genotype frequencies:

	AA	*Aa*	*aa*
BB	$(0.25)(0.64) = 0.16$	$(0.5)(0.64) = 0.32$	$(0.25)(0.64) = 0.16$
Bb	$(0.25)(0.32) = 0.08$	$(0.5)(0.32) = 0.16$	$(0.25)(0.32) = 0.08$
bb	$(0.25)(0.04) = 0.01$	$(0.5)(0.04) = 0.02$	$(0.25)(0.04) = 0.01$

d. Linkage affects only the time to equilibrium, not the equilibrium itself. So there is no effect of $r = 0.001$ upon the equilibrium genotype frequencies.

12.

Genotype	*MM*	*MN*	*NN*
a. Frequency	$61/152 = 0.4013$	0.4211	0.1776
c. Hardy–Weinberg	$(0.612)^2 = 0.3744$	$2(0.612)(0.388) = 0.475$	$(0.388)^2 = 0.1507$
d. Expectation	$0.3744(152) = 560.901$	72.197	22.901
$(o - e)^2/e$	0.2952	0.9307	0.7335
Sum = 1.9595,	df = 1,	$p = 0.16$	

Gametes	M	N
b. 4013 + 0.5(0.4211) =	0.612	0.388

Fail to reject Hardy–Weinberg for this population at the 5% level.

13. **a.** $p^2 = 0.4013$ (the non-N phenotype = MM), so $p = 0.6335$

b. $2pq = 0.4644$. Ratio $= 2pq/p^2 = 1.1572 : 1$

14. **a.** Frequency of MD in European Americans $= (0.24)(0.6) = 0.144$
Frequency of MD in African Americans $= (0.09)(0.84) = 0.0756$

b. Frequency of MD in total population $= 0.7(0.144) + (0.3)(0.0756) = 0.1235$.
Frequency of M in total population $= 0.7(0.24) + (0.3)(0.09) = 0.195$
Frequency of D in total population $= 0.7(0.6) + (0.3)(0.84) = 0.672$

From equation 2.9, $D = 0.1235 - (0.195)(0.672) = -0.0075$. The town population has linkage disequilibrium.

c. Frequency of $Rh+$ within European Americans $= 1 - (1 - p)^2 = 1 - 0.4^2 = 0.84$
Frequency of $Rh+$ within African Americans $= 1-(1-p)^2 = 1-0.16^2 = 0.9744$
Frequency of $Rh+$ in town $= 0.7(0.84) + (0.3)(0.9744) = 0.8803$
Frequency of D in total population $= 0.672$, so frequency of $Rh+$ in town with Hardy–Weinberg is $1 - (1 - p)^2 = 1 - 0.328^2 = 0.8924$
There are fewer $Rh+$'s than expected under Hardy–Weinberg.

d. Expected under random mating $= 0.8924$ (see above)
$\mathrm{D}_1 = (1 - r)\mathrm{D}_0 = 0.5(-0.0075) = -0.0038$

15. The matings at risk are (mother × father) $dd \times DD$ and $dd \times Dd$. Under random mating, the probability of an at-risk mating is (see Table 2.3), in European Americans,

$$(0.4)^2 \times (0.6)^2 + (0.4)^2 \times [2(0.4)(0.6)] = 0.1344$$

Under random mating, the probability of an at-risk mating is, in African Americans,

$$(0.16)^2 \times (0.84)^2 + (0.16)^2 \times [2(0.16)(0.84)] = 0.0249$$

16. a. $D = (0.4)(0.1) - (0.3)(0.2) = -0.02$
$p_A = 0.4 + 0.3 = 0.7$; $p_a = 0.2 + 0.1 = 0.3$; $p_B = 0.4 + 0.2 = 0.6$;
$p_b = 0.3 + 0.1 = 0.4$
$p_A p_B = (0.7)(0.6) = 0.42$; $p_a p_b = (0.3)(0.4) = 0.12$
$D' = -0.02/(0.12) = -0.1667$

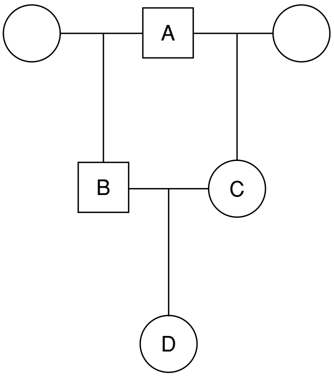
b. $D_1 = (1 - r)D_0 = 0.8(-0.02) = -0.016$; $D_2 = 0.8(-0.016) = -0.0128$

c. $D_1 = (1 - r)D_0 = 0.99(-0.02) = -0.0198$; $D_2 = 0.99(-0.0198) = -0.0196$

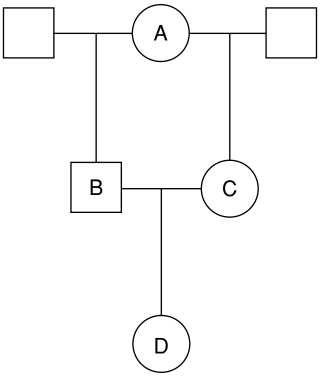
CHAPTER 3

Problems

1. Consider the following two pedigrees:



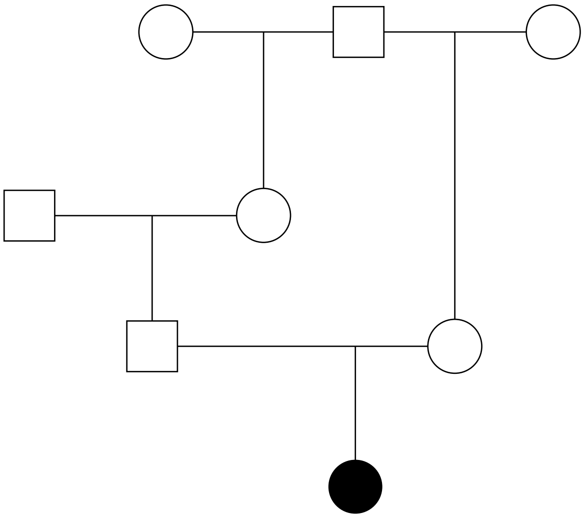
a.



b.

What is the probability of identity by descent in individual D for a randomly chosen X-linked locus in pedigrees a and b?

2. Consider the pedigree below:



What is the pedigree inbreeding coefficient F of the female shown by the solid circle under the assumption that all relevant genetic relationships are shown in the pedigree.

3. Given the following allele frequencies for an autosomal locus with two alleles (A and a, with p being the frequency of A) and inbreeding coefficients (measured as a deviation from Hardy–Weinberg proportions in all problems in this set), calculate the genotype frequencies.

a. $p = 0.1$, $f = 0.5$

b. $p = 0.8$, $f = 0.1$

c. $p = 0.3$, $f = 0.05$

d. $p = 0.3$, $f = -0.05$

4. Estimate the value of f for each population given the following genotype frequencies.

	GENOTYPE		
	AA	Aa	aa
a.	0.30	0.40	0.30
b.	0.20	0.60	0.20
c.	0.056	0.288	0.656
d.	0.024	0.352	0.624

5. Estimate the value of f for each population given the following genotype numbers.

	GENOTYPE		
	AA	Aa	aa
a.	60	0	100
b.	20	60	20
c.	50	50	50

Pool all of these populations together and calculate the value of f for the combined population.

6. Consider the model of assortative mating shown in Figure 3.6. Calculate the genotype and allele frequencies for the first three generations of assortative mating starting with a population in Hardy–Weinberg equilibrium and $p = 0.6$.

7. Consider the model of disassortative mating shown in Figure 3.8. Calculate the genotype and allele frequencies for the first three generations of disassortative mating starting with a population in Hardy–Weinberg equilibrium and $p = 0.6$.

8. The following data were collected on species X at two locations. At this autosomal locus, there are two alleles, A and B.

Location	AA	AB	BB
1	144	471	384
2	64	32	4

(a) For each population, calculate f, the system of mating "inbreeding" coefficient that measures deviations from random mating.

(b) Pool the two populations together and estimate f for the pooled population. Is the deviation from random-mating expectations reduced or intensified by this pooling?

9. Suppose a deme is scored for three randomly chosen loci. The genotypic frequency results are as follows:

Locus i	A_iA_i	A_ia_i	a_ia_i
1	0.09	0.42	0.49
2	0.36	0.48	0.16
3	0.111	0.378	0.511

What do you conclude about this population's system of mating (assume no selection or drift) and why?

10. Same as problem 9, but now assume the results were:

1	0.09	0.42	0.49
2	0.16	0.48	0.36
3	0.024	0.352	0.624

11. Same as problem 9, but now assume the results were:

1	0.1005	0.3990	0.5005
2	0.1720	0.4560	0.3720
3	0.0480	0.3040	0.6480

12. Consider two loci each with two alleles (A and a; B and b) with recombination frequency of 0.3 between them. Population 1 has the frequency of A as 0.5 and the frequency of

B as 0.1, with no linkage disequilibrium, whereas population 2 has the frequency of A as 0.9 and the frequency of B as 0.9, once again with no linkage disequilibrium. These two populations are brought together to form a new admixed population consisting of 60% population 1 and 40% population 2.

a. What is the linkage disequilibrium in the initial admixed population?

b. Suppose population 1 speaks Frigglepix and population 2 speaks Tharkian and there is 100% assortative mating by language in the admixed population. What is the linkage disequilibrium after one generation of assortative mating?

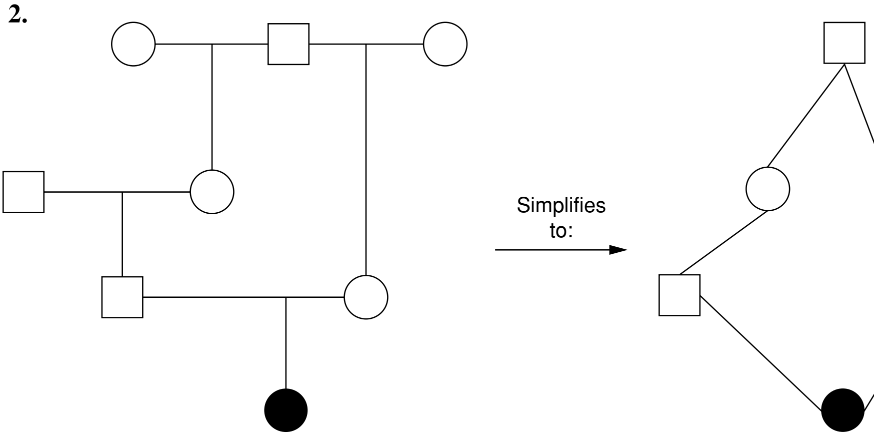
c. Suppose instead that there is random mating in the admixed population. What is the linkage disequilibrium after one generation of random mating?

Answers

1. a. Individual B receives his only X chromosome from his mother, who is not a common ancestor. Therefore, the probability of identity by descent in D is zero.

b. Individual A passes on X chromosomes to both B and C, and the probability of a match is $\frac{1}{2}$. Individual B passes on his X with probability 1 to D, and C passes on her X obtained from A with probability $\frac{1}{2}$. Hence, the total probability of identity by descent is $\frac{1}{4}$.

2.



Calculate F as $\left(\frac{1}{2}\right)^4 = 0.0625$.

3.

	AA	Aa	aa
a. $p = 0.1, f = 0.5$	0.055	0.09	0.855
b. $p = 0.8, f = 0.1$	0.656	0.288	0.056
c. $p = 0.3, f = 0.05$	0.1005	0.399	0.5005
d. $p = 0.3, f = -0.05$	0.0795	0.441	0.4795

4.

$p = G_{AA} + \left(\frac{1}{2}\right)G_{Aa}$	$f = 1 - G_{Aa}/(2pq)$
a. 0.5	0.2
b. 0.5	−0.2
c. 0.2	0.1
d. 0.2	−0.1

5.

	GENOTYPE		
AA	*Aa*	*aa*	
a. 60	0	100	$f = 1 = 1-$ observed heterozygote/(expected $= 2pq$)
b. 20	60	20	$f = -0.2$
c. 50	50	50	$f = \frac{1}{3} = 0.3333$

Pool all of these populations together and calculate the value of f for the combined population.

	AA	*Aa*	*aa*	
Pooled population	130	110	170	$f = 0.4583$

6.

Generation	*AA*	*Aa*	*aa*	*A*	*a*
0	0.36	0.48	0.16	0.6	0.4
1	0.48	0.24	0.28	0.6	0.4
2	0.54	0.12	0.34	0.6	0.4
3	0.57	0.06	0.37	0.6	0.4

7.

Generation	*AA*	*Aa*	*aa*	*A*	*a*
0	0.36	0.48	0.16	0.6	0.4
1	0.28125	0.59375	0.125	0.578125	0.421875
2	0.30212	0.56360	0.13428	0.583922	0.416078
3	0.29714	0.57079	0.13206	0.582540	0.417460

8. a. First, calculate the observed genotype frequencies. Then, using the allele frequencies calculated from the genotype frequencies, determine the expected Hardy–Weinberg heterozygote frequencie. For example, for locality 1:

AA	*AB*	*BB*
144/999	471/999	384/999
0.144	0.471	0.384

Frequency of $A = p = 0.144 + 0.5(0.471) = 0.38$

Frequency of $B = q = 0.62$

Expected heterozygotes $= 2pq = 0.471$

$$f = 1 - \frac{\text{observed heterozygotes}}{\text{expected heterozygotes}} = 1 - \frac{0.471}{0.471} = 0.0$$

If you follow these steps for the other population, you will find $p = 0.8$ and $f = 0.0$.

b. $p = 0.4181$ and $f = 0.0594$. The pooling of the populations has created a heterozygote deficiency from Hardy–Weinberg.

9. $f = 0$ for $i = 1, 2$; $f = 0.1$ for $i = 3$. Implies random mating at loci 1 and 2; assortative mating at locus 3. It cannot be inbreeding because inbreeding affects all loci simultaneously.

10. $f = 0$ for $i = 1, 2$; $f = -0.1$ for $i = 3$. Implies random mating at loci 1 and 2; disassortative mating at locus 3.

11. $f = 0.05$ for all i. Implies inbreeding, not assortative mating given three "randomly chosen loci."

12. **a.** What is the linkage disequilibrium in the initial admixed population? Use $D_{\text{admixture}} = m(1-m)(p_1-p_2)(k_1-k_2)$, $m = 0.6$, $p_1 = 0.5$, $p_2 = 0.9$, $k_1 = 0.1$, $k_2 = 0.9$, so $D = 0.0768$.

b. Same as previous generation, so $D = 0.0768$.

c. $D_1 = (0.0768)(1-0.3) = 0.0538$

CHAPTER 4

Problems

1. An allele has a frequency of 0.01 in a population. If genetic drift is the only force operating, what is the probability under random mating that the allele will ultimately be lost if the population size is 50? Repeat the calculation for a population size of 5,000,000. Repeat both calculations under the assumption that the population is now inbreeding with $f = 0.1$.

2. One hundred populations of size 10 are all started with two alleles (A and a) at a locus, with the A allele having a frequency of 0.2. Eventually, all populations are fixed for one allele or the other, with 32 being fixed for A. Test the hypothesis (using a chi-square statistic, stating degrees of freedom) that this result is consistent with genetic drift being the sole evolutionary force operating on this locus.

3. Eight thousand isolated populations of size 4, each with the same initial frequency of allele A (p), are simulated with the following results:

NUMBER OF ISOLATES WITH y A ALLELES

Generation	$y = 0$	1	2	3	4	5	6	7	8
5	5516	537	564	446	335	238	175	105	84
10	6403	179	221	196	174	158	149	119	401
20	6887	39	47	51	33	55	42	42	804
40	7029	3	2	4	4	3	4	4	947
60	7038	0	0	0	0	0	0	0	962

For the questions below, you will need to calculate means and variances of allele frequencies. Let p_y be the frequency of A in an isolate with y copies of A, n_y be the number of simulated populations with allele frequency p_y, and $n = \sum n_y = 8000$ the total number of simulations. Then the mean and variance of allele frequencies are given by

$$\bar{p} = \sum_{y=0}^{8} \frac{n_y p_y}{n} \quad \mathrm{Var}(p) = \sum \frac{n_y (p_y - \bar{p})^2}{n-1}$$

a. What is the average frequency of A in each generation? Does the average allele frequency over all isolates change much with increasing generation time?

b. What is the most likely original allele frequency?

c. What is the variance in allele frequency among isolates at generations 5, 20, and 60?

d. What proportion of the demes have fixation or loss of A at generations 5, 20, and 60?

e. What evolutionary force is causing these simulated populations to evolve? Justify your answer in terms of the results obtained in parts a, c, and d.

4. Ten replicates of a population evolving at a single locus with two alleles (A and a) are simulated in a computer with the following results for the A allele frequencies:

Trial	Generation 2	Generation 25	Generation 50
1	0.53	0.56	0
2	0.55	0.15	1
3	0.4	0.52	0
4	0.45	0.5	0.67
5	0.48	0.43	0.3
6	0.5	0.61	0.83
7	0.47	0.42	0.67
8	0.48	0.32	0.59
9	0.53	0.36	1
10	0.55	0.7	0.35

What is the variance effective size of the populations at generations 2, 25, and 50 relative to generation 0 (the initial generation at which all replicates had identical gene pools)? Is there any evidence for a bottleneck, and if so, between what generations did it occur?

5. a. Calculate $\bar{F}(1)$ through $\bar{F}(5)$ for populations of sizes $N = 2$, 10, 50 from equation 4.3. Assume $\bar{F}(0) = 0$.

b. Calculate the average inbreeding coefficients for a population of size 50 at generations 0, 1, and 2 that is decreased to size 2 at generations 3, 4, and 5. What is the inbreeding effective size of this population at generation 5 relative to generation 0?

6. Five isolated demes are drawn at random from zygotes produced by 10,000 unrelated adults mating at random for a locus with an allele having an initial frequency of 0.5. Subsequently, the allele frequencies in the individuals drawn to form the five demes are measured to be 0.9, 0.7, 0.6, 0.2, and 0.1. Assume all demes are of equal size (both census and effective in all senses) and have equivalent population structures and genetic drift is the only evolutionary force operating. Use the population of 10,000 unrelated adults as the reference generation.

a. What is the inbreeding effective size of each deme?

b. What is the variance in observed allele frequencies? What is the variance effective size of each deme?

7. a. A captive population of an endangered species is started at a zoo with eight founding individuals. After five generations in captivity, the average F is found to be 0.1. Using the founders as the reference generation, what is the inbreeding effective size over this time period?

b. Assume that a second, isolated population was started with 12 founders at another zoo. After five generations in captivity, this population also had an average F of 0.1.

Using its 12 founders as the reference generation, what is the inbreeding effective size over this time period?

8. Ten unrelated, hermaphroditic individuals from an idealized population produce 100 progeny under random mating.
   a. What are the inbreeding effective and variance effect sizes of the progeny generation relative to the parental generation?
   b. Suppose the parental population consisted of 10 full sibs from a mated pair of two unrelated individuals (the grandparents of the progeny generation). What is the inbreeding effective size of the 100 progeny relative to their grandparental generation?

9. What are the relative sizes of the inbreeding and variance effective sizes given the following demographic events?
   a. A founder event, using the large population the generation before the founder event as the reference generation and the founder generation itself for calculating effective sizes
   b. A founder event, using the founder generation as the reference generation and calculating effective sizes 10 generations later, during which time the population grew at a rapid rate
   c. A large idealized population of 10,000 with the sole exception that $f = 0.1$.
   d. A steadily declining population size

10. Five isolated demes are drawn at random from zygotes produced by 5000 unrelated adults mating at random (the reference population) for a locus with an allele having a frequency of 0.5 in the reference population. Subsequently, the allele frequencies in the individuals drawn to form the five demes are measured to be 0.7, 0.6, 0.5, 0.4, and 0.3. Assume all demes are of equal size (both census and effective in all senses) and have equivalent population structures and genetic drift is the only evolutionary force operating.
   a. What is the inbreeding effective size of each deme?
   b. What is the variance effective size of each deme?

Answers

1. The probability of ultimate loss = 1− Prob(ultimate fixation)= $1 - p_0 = 0.99$. This is true for all population sizes and values of f.

2. The probability of fixation of A under drift = $p_0 = 0.2$. So the expected number of populations fixed for A under drift = 100(0.2) = 20 and the expected number fixed for a is 100(0.8) = 80. The observed numbers are 32 fixed for A and 68 fixed for a. The chi-square statistic is therefore $(32 - 20)^2/20 + (68 - 80)^2/80 = 9$. There is one degree of freedom, so we reject the null hypothesis of drift with a p value of 0.0027.

3.

Generation	$\bar{p}$	Var(p)	Proportion fixed
5	0.125	0.0527	(5516 + 84)/8000 = 0.70
10	0.119		
20	0.120	0.0982	(6887 + 804)/8000 = 0.96
40	0.120		
60	0.120	0.1058	1

a. See the table above; $\bar{p}$ does not change much.

b. Since $2N = 8$, the most likely $p_o = 0.125\,(= \frac{1}{8})$ because $\bar{p} = 0.125$ at generation 5, the generation closest to the initial conditions and therefore the most likely to be closest to the initial conditions.

c. See the table above.

d. See the table above.

e. Genetic drift because $\bar{p} \approx$ constant, var(p) increases with time, and the number of fixed demes increases with time.

4.

	ALLELE FREQUENCY		
Trial/Generation	2	25	50
1	0.53	0.56	0
2	0.55	0.15	1
3	0.4	0.52	0
4	0.45	0.5	0.67
5	0.48	0.43	0.3
6	0.5	0.61	0.83
7	0.47	0.42	0.67
8	0.48	0.32	0.59
9	0.53	0.36	1
10	0.55	0.7	0.35
Average p	0.494	0.457	0.541
Variance p	0.002293333	0.024823333	0.13561
N_{ev}	54.49796512	114.1050038	31.26706232

Use

$$N_{ev} = \frac{1}{2\left\{1 - \left[1 - \sigma_t^2/(pq)\right]^{1/t}\right\}}$$

The results are given in the table above. As can be seen, the effective size drops between generations 25 and 50, indicating a bottleneck some time in that period.

5. a. Below are the values for $\bar{F}(i)$:

	POPULATION SIZE		
Generation	2	10	50
0	0	0	0
1	0.25	0.05	0.01
2	0.4375	0.0975	0.0199
3	0.578125	0.142625	0.029701
4	0.68359375	0.18549375	0.03940399
5	0.762695313	0.226219063	0.04900995

b. $\bar{F}(i)$ depends on the number of parents at $i - 1$. For the individuals produced at generation 3, there are still 50 parents from the previous generation. Hence,

Generation	$\bar{F}(i)$	Number of Parents
0	0	50
1	0.01	50
2	0.0199	50
3	0.029701	50
4	0.27227575	2
5	0.454206813	2

Use

$$N_{ef} = \frac{1}{2\left\{1 - \left[1 - \bar{F}(t)\right]^{1/t}\right\}}$$

with $\bar{F}(5) = 0.45420681$, $t = 5$, to yield $N_{ef} = 4.3838$.

6. a. $N_{ef} = 10{,}000$ (the number of parents)

b. $\sigma_p^2 = 0.2\sum(p_i - 0.5)^2 = 0.092$
$p_o q_o/(2N_{ev}) = 0.25/(2N_{ev}) = 0.092$
$N_{ev} = 0.125/0.092 = 1.36$

7. a. Use

$$N_{ef} = \frac{1}{2\left\{1 - \left[1 - \bar{F}(t)\right]^{1/t}\right\}}$$

with $t = 5$ and $\bar{F} = 0.1$ to yield $N_{ef} = 24$.

b. Nothing has changed: $t = 5$ and $\bar{F} = 0.1$ to yield $N_{ef} = 24$.

8. a. Since the population is ideal and $\bar{F}(0) = 0$, N_{ef} = number of parents = 10 and N_{ev} = number of offspring = 100.

b. The $\overline{F}$ (offspring generation) $= \frac{1}{20} + (1 - \frac{1}{20})(0.25)$ [the probability of selfing times $\frac{1}{2}$ (the inbreeding coefficient of offspring from a self mating) + the probability of not inbred through selfing times $\frac{1}{4}$ (the inbreeding coefficient for offspring of full sibs)] = 0.2875. Use

$$N_{ef} = \frac{1}{2\left\{1 - \left[1 - \bar{F}(t)\right]^{1/t}\right\}}$$

with $t = 2$ (two generations from grandparents) and $\bar{F} = 0.2875$ to yield $N_{ef} = 3.21$.

9. a. For the founder generation N_{ef} is minimally affected because the number of parents in the reference generation is large; N_{ev} is strongly affected and will be very small because the number of offspring (the founder generation) is small.

b. Here N_{ef} is very small because of the founder event 10 generations ago (the number of ancestors is small); N_{ev} is larger than N_{ef} because the population size is growing, reducing the variance in allele frequency due to less drift with increasing sample size.

c. Here N_{ef} is small because identity by descent accumulates rapidly due to the system of mating; N_{ev} is large because it is not as sensitive to f as N_{ef}.

d. Here N_{ef} will be larger than N_{ev} because the number of parents is larger than the number of offspring.

10. a. The inbreeding effective size in this case depends only upon the number of unrelated parents, which is 5000.

b. First, calculate the variance of allele frequencies across the replicates: $\sigma_p^2 = \sum 1/5(p_i - 0.5)^2 = 0.02$.

Second, from the fundamental definition of variance effective size, we can solve for N$_{ev}$: $p_o q_o/(2N_{ev}) = 0.25/(2N_{ev}) = 0.02$. So $N_{ev} = 0.125/0.02 = 6.25$.

CHAPTER 5

Problems

1. Let $\mu = 10^{-6}$ be the neutral mutation rate at a locus. Assuming all alleles are neutral and mating is at random, what is the expected average equilibrium probability of identity by descent at this locus for inbreeding effective sizes of 1000, 10,000 and 100,000?

2. a. A population of size 1000 is ideal except that the number of offspring per individual has a mean of 2 and a variance of 4. Assume that the neutral mutation rate is 10^{-6} for a particular locus. What is the rate of neutral molecular evolution at this locus? What is the expected average inbreeding coefficient in this population at equilibrium between mutation and drift? What is the expected equilibrium heterozygosity for neutral alleles at this locus in this population?

b. Assume the population size is changed to 100,000. Redo all the calculations made in part a.

c. Assume that the population size is still 100,00 but that the neutral mutation rate is now 10^{-5}. Redo all the calculations made in part a.

3. Two homologous, autosomal genes are sampled from a population with an inbreeding effective size of 100. What is the probability that these two genes coalesced in the previous generation? What is the probability that they coalesced 100 generations ago?

4. Two homologous, autosomal genes are sampled from a population with an inbreeding effective size of 1000 with the mutation rate at the locus being 10^{-5} per generation.

a. What is the probability that these two genes coalesced before mutation 500 generations ago?

b. What is the probability that one of these gene lineages experienced mutation before coalescence 500 generations ago?

c. What is the probability of mutation before coalescence given that either mutation or coalescence occurred 500 generations ago?

d. What is the expected heterozygosity for this system under the neutral theory?

5. **a.** A population of size 100 is ideal except for a deviation from random mating of $f = 0.1$. Assume that variance effective size and eigenvalue effective size are identical in this population. Assume that the neutral mutation rate is 10^{-7} for a particular locus. What is the rate of neutral molecular evolution at this locus?

   b. Assume the population size is changed to 1000. What is the rate of neutral molecular evolution at this locus?

6. A DNA region of nine nucleotides is sequenced, with variable sites being found that define five haplotypes as follows:

   1. A C C G T T G C A
   2. T C G G T C G C A
   3. A C G G T T G C A
   4. A C C G T T G C C
   5. A C G G T C G T A

   Construct the maximum-parsimony haplotype network for these sequences assuming no recombination and no multiple mutation events. Can you infer any haplotypes that must have existed but are not present in this sample? What is/are the sequence(s) of such haplotype(s)?

7. **a.** Two autosomal genes are sampled at random from a deme of inbreeding effective size 500 and variance effective size 1000. What are the mean and variance to the time of coalescence of these two genes?

   b. Assume that 10 genes are sampled. What are the mean and variance to the first coalescent event involving these 10 genes?

   c. Assume again that 10 genes are sampled. What are the mean and variance to the time of coalescence for all 10 genes?

8. Some VNTR loci (Appendix 1) have a mutation rate of 10^{-2}. In a population of inbreeding effective size of 25, calculate the probability of coalescence for two randomly drawn alleles for one, two, and three generations in the past. Calculate the probability of coalescence before mutation and the probability of mutation before coalescence for two randomly drawn alleles. What is the expected heterozygosity at this VNTR locus in this population if it were randomly mating for this locus?

Answers

1. $F_{\text{eq}} = 1/[4N\mu + 1]$, $\mu = 10^{-6}$, so:

N	F_{eq}
10^3	0.996
10^4	0.9615
10^5	0.7143

2. **a.** The rate of neutral molecular evolution = the neutral mutation rate = 10^{-6}. Use equation 4.31 with $k = 2$ and $v = 4$ to get $N_{ef} = 666.78$. From equation 5.6, $F_{\text{eq}} = 0.9973$ and heterozygosity = 0.0027.

b. The rate of neutral molecular evolution = the neutral mutation rate = 10^{-6}:

N	100,000
N_{ef}	66,666.77778
F	0.789473407
H	0.210526593

c. The rate of neutral molecular evolution = the neutral mutation rate = 10^{-5}:

N	100,000
N_{ef}	66,666.77778
F	0.272726942
H	0.727273058

3. Use Prob(coalesce at t) = $[1 - 1/(xN_{ef})]^{t-1}[1/(xN_{ef})]$ with $N_{ef} = 100$, $x = 2$, and $t = 1$ to yield 0.005 and with $t = 100$ to yield 0.00304.

4. a. Use equation 5.11 to get the probability of coalescence before mutation to be 0.000385695.

b. Use equation 5.12 to get the probability of mutation before coalescence to be 1.54×10^{-5}.

c. Probability of mutation before coalescence given one has occurred is $1.54 \times 10^{-5}/(1.54 \times 10^{-5} + 0.000385695) = 0.038443417$.

d. Use equation 5.14 or 5.7 to get expected heterozygosity. First, $\theta = 4 \times 1000 \times 10^{-5} = 0.04$. Expected heterozygosity is $\theta/(1 + \theta) = 0.038452$.

5. a. Rate of neutral evolution = neutral mutation rate = 10^{-7}.

b. Rate of neutral evolution = neutral mutation rate = 10^{-7}.

6. The tree is 4 ↔ 1 ↔ 3 ↔ 0 ↔ 2 where arrows indicate single nucleotide changes
↓ and zero an intermediate haplotype not
5 present in the sample.

Yes, and the sequence must be A C G G T C G C A.

7. a. Expected coalescence time = $2N_{ef} = 1000$ generations. Var = $2N_{ef}(2N_{ef} - 1) = 1000(999) = 999{,}000$.

b. For $n = 10$, the expected time for first coalescence is $4N_{ef}/(10 \times 9) = 44.44$ generations.

$$\text{Var} = \frac{4N_{ef}}{10 \times 9}\left(\frac{4N_{ef}}{10 \times 9} - 1\right) = 471.60$$

c. For $n = 10$, the expected coalescence time = $4N_{ef}(1 - 1/n) = 2000(1 - 0.1) = 1800$ generations.

$$\text{Var} \approx 16N_{ef}^2 \sum_{i=2}^{10} \frac{1}{(i)^2(i-1)^2} = 2316.28$$

8. Prob(coalescence at generation 1) $= 1/(2N_{ef}) = 1/50 = 0.02$
Prob(coalescence at generation 2) $= [1 - 1/(2N_{ef})][1/(2N_{ef})] = 0.01960$
Prob(coalescence at generation 3) $= [1 - 1/(2N_{ef})]^2[1/(2N_{ef})] = 0.01921$

$\theta = 4N_{ef}\mu = 1$
Prob(mutation before coalescence) $= \theta/(1+\theta) = 0.5$
Prob(coalescence before mutation) $= 0.5$
Expected heterozygosity $= \theta/(1+\theta) = 0.5$

CHAPTER 6

Problems

1. Suppose that the neutral allele A has a frequency of 0.4 in population 1 and 0.7 in population 2. After a single generation of gene flow between these populations, the frequency of A in population 1 is 0.46. What proportion of gametes (m) entered population 1 from population 2 to explain this result (assume that the two populations are effectively infinite in size).

2. Two populations exchange gametes at a rate of 0.01 every generation. Given that the two populations were initially fixed for different alleles at an autosomal locus, what is the expected difference (assuming neutrality and infinite population size in each deme) in the allele frequency between the demes at generation 10 and at generation 100? Now assume that the rate of exchange is 0.001. Redo the calculations for generations 10 and 100. What are the expected differences in allele frequency at equilibrium for $m = 0.01$ and $m = 0.001$?

3. Two autosomal genes picked at random from within a randomly chosen local deme coalesce on the average at 150 generations ago. Two autosomal genes picked at random from the entire species coalesce on the average at 1000 generations ago. What is the F_{st} for this species?

4. Three populations are screened for genetic variability at an autosomal locus with two alleles, A and a, with the following results:

	AA	Aa	aa
Population 1	12	176	312
Population 2	90	220	90
Population 3	469	462	69

Calculate f_{is}, f_{st}, and f_{it}.

5. A species fits the one-dimensional stepping-stone model with each deme having a variance effective size of 50 and with $m_1 = 0.1$ and $m_\infty = 0.001$.

a. What is the equilibrium f_{st} for this species?

b. Double the amount of local gene flow and recalculate f_{st}.

c. Go back to the original parameters, but now double the long-distance gene flow parameter and recalculate f_{st}. Which parameter doubling had the greatest effect on f_{st}? Which doubling has the greatest effect on the number of gametes being exchanged between demes?

6. Suppose that the allele A is fixed in population 1 and absent in population 2. Now assume that the two populations exchange 10% of their genes every generation and each population is effectively infinite in size.

a. What are the allele frequencies in each population after one and after two generations of gene flow?

b. What will the allele frequency of A be in populations 1 and 2 after a large number of generations of gene flow?

7. Four populations are screened for genetic variability at an autosomal locus with two alleles, A and a, with the following results:

Population	Population 1			Population 2			Population 3			Population 4		
Genotype	*AA*	*Aa*	*aa*	*AA*	*Aa*	*aa*	*AA*	*Aa*	*aa*	*AA*	*Aa*	*aa*
Frequency	0.019	0.162	0.819	0.184	0.432	0.384	0.275	0.45	0.275	0.511	0.378	0.111
Population size	2000			3000			3000			2000		

a. Calculate f_{is}, f_{st}, and f_{it}.

b. Calculate the effective number of migrating individuals among these populations under an island model. Is this a variance or inbreeding effective number of migrants and why?

8. A population has a 50–50 sex ratio, and males disperse twice as much as females (for females $m = 0.001$), all according to an island model and in which dispersal equals gene flow. All local populations are "ideal" in an effective size sense and each has a size of 1000.

a. What is the expected F_{st} for an autosomal locus?

b. What is the expected F_{st} for mtDNA (assume standard maternal, haploid inheritance)?

c. What is the expected F_{st} for Y-DNA (assume a standard XY sytem of sex determination)?

9. Several replicates are taken from a reference population such that each replicate has a variance effective size of 10. An autosomal locus with two alleles has an allele frequency of 0.8 in the reference population.

a. What is the variance in allele frequency among the replicates at this locus?

b. What value of F_{st} do you expect among the replicates?

Imagine that all replicates are carried on for two more generations (three generations from the ancestral reference population) and that the variance effective size is maintained to be a constant 10 each generation. At this third generation:

c. What is the variance in allele frequency among the replicates at this locus?

d. What value of F_{st} do you expect among the replicates?

10. The following allele frequencies for the *M–N* blood group locus were estimated for these populations:

Population	West Africa	African Americans, Georgia	European Americans, Georgia
Frequency of *M* allele	0.476	0.484	0.507

Calculate *M*, the net amount of gene flow from the Georgia European American population into the Georgia African American population (assume gene flow has been essentially a one-way phenomenon from European Americans to African Americans).

11. A species is subdivided into four subpopulations with the following genotype frequencies at a locus:

	AA	*Aa*	*aa*
Population 1	0.1005	0.3990	0.5005
Population 2	0.1720	0.4560	0.3720
Population 3	0.0480	0.3040	0.6480
Population 4	0.8145	0.1710	0.0145

Assume all populations are of equal size. Calculate f_{is}, f_{st}, and f_{it}.

12. A population is continuously distributed over a one-dimensional habitat with a density of two individuals per meter. The standard deviation (i.e., the square root of the variance) between place of birth and place of reproduction is 4 meters. If 0.1% of the individuals engage in long-distance dispersal at random over the entire population, what is the expected equilibrium value of f_{st}? Redo the calculations assuming 1% of the individuals engage in long-distance dispersal.

Now suppose that all is the same as above, except that the habitat is two dimensional and the density is $2^2 = 4$ individuals per square meter. Calculate the equilibrium f_{st} values of 0.1% and 1% long-distance dispersal.

Answers

1. Use equation 6.1 and solve for m : $p_1' = (1 - m)p_1 + mp_2$ or $0.46 = (1 - m)0.4 + m0.7, m = 0.2$.

2. Use equation 6.8 with $d_0 = 1, d_t = (1 - 2m)^t$. It is easiest to use logarithms; $\ln d_t = t\ln(1 - 2m)$.
For $m = 0.01$, $\ln d_{10} = 10(\ln 0.98) = -0.202$, so $d_{10} = 0.8171$. At $t = 100$, $d_{10} = 0.1326$.
For $m = 0.001$, $\ln d_{10} = 10(\ln 0.998)$, so $d_{10} = 0.9802$. At $t = 100$, $d_{10} = 0.8186$.
For both m's, the equilibrium difference is zero.

3. Use equation 6.14: $F_{st} = (1000 - 150)/1000 = 0.85$.

4. Convert from numbers to frequencies by dividing by the population size, and also divide each population size by the sum of all population sizes to get the w_i's. Then calculate the average allele frequency (0.5263) and the variance in allele frequency (0.0440 from equation 6.22) to yield $f_{st} = 0.1767$. Within each population, calculate f (equation 3.3) to get -0.1 in each case, so $f_{is} = -0.1$. Apply equation 3.3 to the genotype numbers in the total population (the sum of all three local populations) or use the definition of f_{it} below equation 6.30 to get $f_{it} = 0.0943$.

5. a. Use equation 6.34 to get $f_{st} = 0.2612$.

b. When $m_1 = 0.2$, using equation 6.34 now yields $f_{st} = 0.2$.

c. With $m_\infty = 0.002$, using equation 6.34 now yields $f_{st} = 0.2$. Hence, the two doublings have identical effects on f_{st}. However, the doubling of m_1 increases the number of gametes being exchanged much more than the doubling of m_∞.

6. a. $p_1(t = 1) = (0.9)(1) + (0.1)(0) = 0.9$, $p_2(t = 1) = (0.9)(0) + (0.1)(1) = 0.1$
$p_1(t = 2) = (0.9)(.9) + (0.1)(1) = 0.82$, $p_2(t = 2) = (0.9)(0.1) + (0.1)(0.9) = 0.18$

b. $p_1(t = \infty) = p_2(t = \infty) = 0.5$

7. a.

Population	1	2	3	4
f_{is}	$1 - 0.162/.18 = 0.1$	0.1	0.1	0.1

f_{st}: Average $p = \bar{p} = (2000/10{,}000)(0.1) + (0.3)(0.4) + (0.3)(0.5) + (0.2)(0.7) = 0.43$

$\mathrm{Var}(p) = (0.2)(0.1 - 0.43)^2 + (0.3)(0.4 - 0.43)^2 + (0.3)(0.5 - 0.43)^2 + (0.2)(0.7 - 0.43)^2 = 0.0381$

$\bar{p}\bar{q} = 0.2451$

So $f_{st} = 0.0381/0.2451 = 0.155$.

$f_{it} = f_{st} + f_{is}(1 - f_{st}) = 0.24$

Alternative. average frequency of Aa in total population $= 0.373$, so $f_{it} = 1 - 0.373/[2(0.43)(0.57)] = 0.24$.

b. Under an island model, $f_{st} = 1/(1 + 4Nm)$, so $Nm = (1/f_{st} - 1)/4 = 1.36$. This is a variance effective number of migrants because the genetic parameter we are using is the variance of allele frequency.

8. a. For an autosomal locus, $m = (0.001 + 0.002)/2$.
$F_{st} = 1/(1 + 4Nm) = 1/[1 + 4(1000)(0.0015)] = 0.143$

b. For mtDNA, $m = 0.001$ (female only) and $N = 500$ (given the 50–50 sex ratio). Because mtDNA is haploid:

$$F_{st} = 1/(1 + 2Nm) = 1/[1 + 2(500)(0.001)] = 0.5$$

c. For Y-DNA, $m = 0.002$ (male only) and $N = 500$. Because it is haploid:

$$F_{st} = 1/(1 + 2Nm) = 1/[1 + 2(500)(0.002)] = 0.333$$

9. a. The expected variance in allele frequency is $pq/(2N_{ev}) = (0.8)(0.2)/(20) = 0.008$.

b. $f_{st} = \sigma_p^2/(pq) = 0.008/(0.8)(0.2) = 0.05$

c. Var(allele freq. after t generations) $= pq\{1 - [1 - 1/(2N_{ev})]^t\}$, so

$$\text{Var}(t) = (0.8)(0.2)\{1 - [1 - 1/20]^3\} = 0.023$$

d. $f_{st} = \sigma_p^2/(pq) = 0.023/(0.8)(0.2) = 0.143$

10. $M = (p_B - p_{Af})/(p_W - p_{Af}) = 0.258$

11.

Population	$f = f_{is} = 1 - \text{freq}_i(Aa)/(2p_i q_i)$	
1	0.05	
2	0.05	
3	0.05	
4	0.05	$f_{is} = 0.05$ for total population

$\text{Var}(p) = 0.0725$; $f_{st} = \text{Var}(p)/(\bar{p}\bar{q}) = 0.293$
$f_{it} = f_{st} + f_{is}(1 - f_{st}) = 0.328$

Alternative: calculate f_{st} as above, but calculate $f_{it} = 1 - \text{freq}(Aa)/(2\bar{p}\bar{q})$ [note freq(Aa), $\bar{p}$, and $\bar{q}$ refer to the *total* population). Then calculate f_{is} from $f_{it} = f_{st} + f_{is}(1 - f_{st})$.

12.

Dimensionality	m_∞	f_{st}
1	0.001	0.411
1	0.01	0.181
2	0.001	0.0038
2	0.01	0.0024

CHAPTER 7

Problems

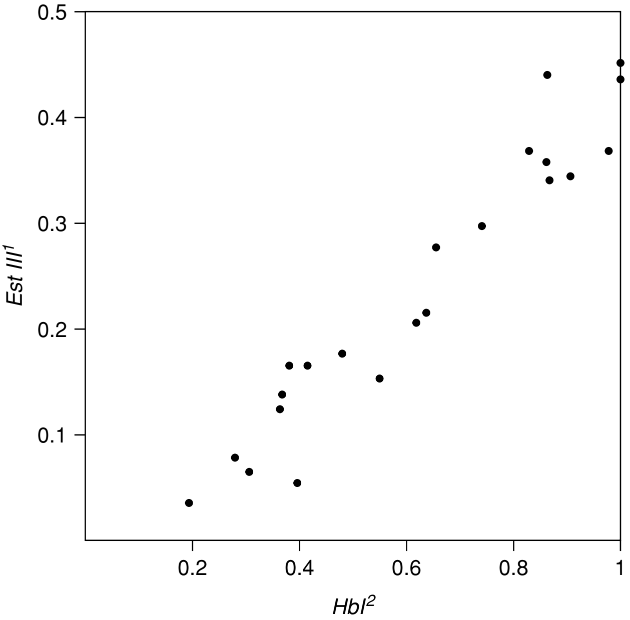
1. A DNA region of nine nucleotides is sequenced, with variable sites being found that define five haplotypes as follows:

1. A C C G T T G C A
2. A C G G T C G C A
3. A C G G T T G C A
4. A C C G T T A C A
5. A C G G T C G T A

a. Construct a cladogram of the haplotypes assuming the least number of mutational events and no recombination.
Suppose that sequence 5 came from another species and can be regarded as a good "outgroup" for the remaining haplotypes.

b. Under an isolation-by-distance model, which haplotype would be expected to have the most restricted geographical distribution and which the broadest?

2. Several Baltic Sea populations of the fish *Zoarces viviparus* were scored for the frequency of the $EstIII^1$ allele at one locus, and the frequency of the HbI^2 allele at another locus with the results shown below:



Draw a rotated set of axes on this diagram that should approximate the results obtained from a principal-component analysis. Identify which of your rotated axes corresponds to the first principal component.

3. A DNA region of 1000 nucleotides is screened with a battery of restriction enzymes, each with a recognition sequence of length 4. A total of 100 chomosomes were screened with the following haplotypes being discovered:
Haplotype (+ = site cut, − = site not cut)

1. + + + − + + − − + + + − − + + − +
2. + + − − + + − − + + + + + + + − +
3. + + + − + + − − + + + − − + + + +
4. + + + − + + − + + + + − − + + − +
5. + + + − + + − − + + + + + + + − +
6. + + + + + + − − + + + − − + + + +
7. + + + − + + − − + + + + − + + − +
8. + + − − + + + − + + + + + + + − +
9. − + + − + + − − + + + − − + + + +

a. Construct a haplotype tree assuming the least number of mutational events and no recombination.

b. Assume that the samples were taken over the entire geographical range of the species. Assume further that all subpopulations within the species are genetically interconnected but that gene flow is restricted by isolation by distance. Assume further that haplotype 1 is closest to the haplotypes found in an outgroup species. Under these assumptions, which haplotypes would be expected to have the more restricted geographical distribution relative to the other and why?

i. 3 vs. 6; ii. 1 vs. 4; iii. 2 vs. 8; iv. 6 vs. 9; v. 4 vs. 8

Which of the following groups (clades) of haplotypes collectively would be expected to have the most widespread geographical distribution and why?

vi. {3,6,9}, {1,4,7}, (5,2,8}

c. Assume that factors other than isolation by distance could be operating and that haplotype 4 has the most widespread geographical distribution and is found in some areas where none of the other haplotypes are found. What does this pattern suggest about the evolutionary history of this species and why?

4. Five loci showing no internal recombination are surveyed in populations spanning the geographical range of species X. Nested-clade analyses of these five loci reveal that each infers a range expansion event into the northern part of its current range from the southern part of its range. A molecular clock is used to time these inferences (t_i) from k_i, the average nucleotide divergence at locus i, with the following results:

Locus	t_i (years before present)	k_i
1	12,000	20
2	20,000	5
3	10,000	10
4	13,000	8
5	21,000	7

a. Assuming all the loci are detecting the same range expansion event, what is the estimate of the time of this range expansion based upon all five loci and what are the variance and standard deviation (square root of the variance) of this estimated time?

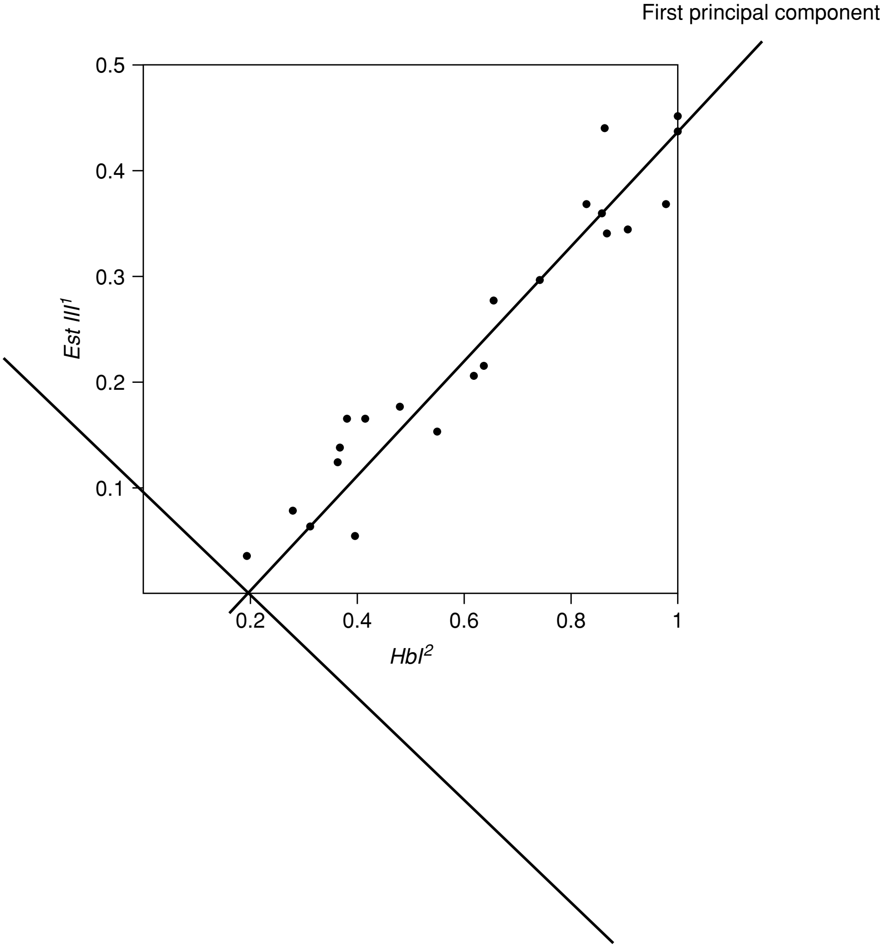
b. Test the assumption that all five loci are detecting the same range expansion event using the 5% level of significance.

Answers

1. a. cladogram: 4–1–3–2–5

b. Since 5 is the outgroup, haplotype 2 is the oldest and 4 the youngest in the ingroup. Under isolation by distance, we expect the geographical range of a haplotype to be positively correlated with its age, so haplotype 4 should have the most restricted geographical range and haplotype 2 the broadest.

2.



3. a. 6—3—1—7—5—2—8
| |
9   4

b. Expect tips to be more restricted than their interiors under isolation by distance. Therefore:

i. 6 (tip) should be more restricted than 3 (its interior);
ii. 4 (tip) should be more restricted than 1 (its interior);
iii. 8 (tip) should be more restricted than 2 (its interior);
iv. no predictions because 6 and 9 are both tips;
v. no predictions because 4 and 8 are both tips.

Let clade $1 - 1 = \{3, 6, 9\}$, clade $1 - 2 = \{1, 4, 7\}$, and clade $1 - 3 = (5, 2, 8\}$, the tree of these clades is $1 - 1 \leftrightarrow 1 - 2 \leftrightarrow 1 - 3$, so $1 - 2$ is the interior clade to the tip clades $1 - 1$ and $1 - 3$. Moreover $1 - 2$ is the oldest clade because it contains the connection to the outgroup. Therefore, expect $1 - 2$ to have the widest distribution.

c. This pattern suggests a range expansion because 4 is the tip to 1, yet 4 is more widespread and found in areas in which the other haploytpes, including 1, are absent. This indicates 4 was in the founding population of the new areas and went to fixation.

4. **a.**

Pooled T (equation 7.9)	13945.45455
Variance of pooled T (equation 7.11)	3825785.124
Standard deviation of pooled T	1955.961432

b. G test statistic (equation. 7.8) $= 12.52$ with four degrees of freedom, which yields a p level of 0.0139. Therefore, we reject the null hypothesis of a single range expansion event.

CHAPTER 8

Problems

1. The *pygmy* locus in mice has two alleles, A and a, that affect body size, with the following genotypic values:

Genotype	Average Body Weight (grams)
AA	14
Aa	12
Aa	6

Suppose a mouse population is randomly mating with $p = 0.6$ (the frequency of A).

a. Calculate the population mean and all genotypic deviations, average excesses, average effects, additive genotypic deviations, nonadditive genotypic deviations, genetic variance, additive genetic variance, and nonadditive genetic variance. Given that the environmental variance is 10 g^2, calculate the broad-sense heritability and the heritability.

b. Redo part a under the assumption that inbreeding, rather than random mating, is occurring with $f = 0.2$ (as a deviation from Hardy–Weinberg genotype frequencies).

2. The following data set is a slightly numerically simplified version of a real data set. A human population was scored for their LDL cholesterol levels and for their genotype at the *ApoE* locus. The *ApoE* locus codes for apoprotein E, which is not found on LDL cholesterol but does competitively bind the LDL receptor protein. There are three common alleles at this locus: $\varepsilon 2$, $\varepsilon 3$, and $\varepsilon 4$. The population was found to be in Hardy–Weinberg equilibrium frequencies at this locus with the following allele frequencies: 0.1 for $\varepsilon 2$, 0.7 for $\varepsilon 3$, and 0.2 for $\varepsilon 4$. The mean phenotypes (measured as mg LDL cholesterol/dl blood plasma) for the various genotypes were as follows:

Genotype	$\varepsilon 2/\varepsilon 2$	$\varepsilon 2/\varepsilon 3$	$\varepsilon 3/\varepsilon 3$	$\varepsilon 2/\varepsilon 4$	$\varepsilon 3/\varepsilon 4$	$\varepsilon 4/\varepsilon 4$
LDL cholesteral	76	90	100	115	110	106

The total phenotypic variance of LDL cholesterol is 554.2 mg^2/dl^2.

a. Calculate the genotypic deviations for all genotypes.

b. Calculate the average excesses and average effects for all alleles.

c. Calculate the breeding values and dominance deviations for all genotypes.

d. Calculate the environmental, genetic, additive genetic, and dominance variances. Give the broad-sense and narrow-sense heritabilities.

e. Increased levels of LDL cholesterol are associated with increased risks of coronary heart disease. Which genotype is at greatest risk? Which genotype has his or her children at greatest risk?

f. Redo parts a through e under the assumption that inbreeding, rather than random mating, is occurring with $f = 0.1$ (as a deviation from Hardy–Weinberg genotype frequencies).

3. Thalassemia is an inherited, autosomal hemoglobin abnormality associated with both anemia and malarial resistance. It is frequent in many human populations in Asia and the Mediterranean where malaria is common. Suppose in Sardinia that the various genotypes have the following average life spans in years:

Genotype	AA	ATh	$ThTh$
Life span	60	65	20

Assume that the standard deviation (square root of the variance) around each of these means is 5 years for each genotypic class (that is, the environmental variance is 25 years^2).

a. What are the genotypic values for the three genotypes?

b. If the allele frequency of Th in Sardinia is 0.1 and the population is in Hardy–Weinberg equilibrium (retain these assumptions in all subsequent questions unless stated otherwise), what is the average life span for an individual from the population?

c. What are the genotypic deviations of life span for the three genotypes?

d. Given a $ThTh$ individual that died at age 27, what is that individual's environmental deviation?

e. Calculate the average excesses for each gamete type and the breeding value for each genotype.

f. Repeat the calculations in parts b, c and e for allele frequencies of 0.2 and 0.05 for Th. At which of these three allele frequencies (0.05, 0.1, 0.2) does Th have the smallest average effect in absolute value? At which allele frequency does the population have the highest average life span?

g. Calculate the narrow-sense heritability of life span in this population at the allele frequencies of 0.05, 0.1, and 0.2. What does the heritability value at the allele frequency of 0.1 tell you about the inheritance of life span in this population?

Answers

1. a.

	GENOTYPE				
	AA	Aa	aa		
Genotype frequency	0.36	0.48	0.16		$f = 0$
Genotypic mean	14	12	6		$p = 0.6$
Genotypic deviation	2.24	0.24	−5.76		Overall mean = 11.76
Average excess	for A:	1.44	for a:	−2.16	
Average effect	for A:	1.44	for a:	−2.16	
Additive deviation	2.88	−0.72	−4.32		

Dominance deviation	−0.64	0.96	−1.44	
Genetic variance	7.1424			
Additive variance	6.2208			
Dominance variance	0.9216			
Heritability				Environmental variance = 10
Broad	0.4167			
Narrow	0.3629			

b.

	GENOTYPE				
	AA	*Aa*	*aa*		
Genotype frequency	0.408	0.384	0.208		$f = 0.2$
Genotypic mean	14	12	6		$p = 0.6$
Genotypic deviation	2.432	0.432	−5.568		Overall mean = 11.568
Average excess	for *A*:	1.792	for *a*:	−2.688	
Average effect	for *A*:	1.4933	for *a*:	−2.24	
Additive deviation	2.9867	−0.7467	−4.48		
Dominance deviation	−0.5547	1.1787	−1.088		
Genetic variance	8.9334				
Additive variance	8.0282				
Dominance variance	0.9052				
Heritability					Environmental variance = 10
Broad	0.4718				
Narrow	0.4240				

2.

	GENOTYPE						
	$\varepsilon2/\varepsilon2$	$\varepsilon2/\varepsilon3$	$\varepsilon3/\varepsilon3$	$\varepsilon2/\varepsilon4$	$\varepsilon3/\varepsilon4$	$\varepsilon4/\varepsilon4$	Variance, Average, or Sum
Hardy–Weinberg genotype	0.010	0.140	0.490	0.040	0.280	0.040	1 (sum)
Genotypic mean	76	90	100	115	110	106	102.0 (ave.)
a. Genotypic deviation	−26.0	−12.0	−2.0	13.0	8.0	4.0	54.2 (var.)
b. Average excess = Average effect	$\varepsilon2$: −8.4		$\varepsilon3$: −1.0		$\varepsilon4$: 7.7		
c. Breeding value	−16.8	−9.4	−2.0	−0.7	6.7	15.4	39.2 (var.)
Dominance deviation	−9.2	−2.6	0.0	13.7	1.3	−11.4	15.0 (var.)

d. From above, $V_g = 54.2$, $V_e = 554.2 - 54.2 = 500$, $V_a = 39.2$, $V_d = 15.0$. Heritability: broad: 54.2/554.2 = 0.098; narrow = 39.2/554.2 = 0.071 .

e. $\varepsilon2/\varepsilon4$ has the highest genotypic value (115), and therefore has the greatest risk. $\varepsilon4/\varepsilon4$ has the highest breeding value (15.4), and therefore its offspring are at greatest risk.

f. for inbreeding with $f = 0.1$:

GENOTYPE

	$\varepsilon2/\varepsilon2$	$\varepsilon2/\varepsilon3$	$\varepsilon3/\varepsilon3$	$\varepsilon2/\varepsilon4$	$\varepsilon3/\varepsilon4$	$\varepsilon4/\varepsilon4$	Variance, Average, or Sum
Genotype frequency	0.019	0.126	0.511	0.036	0.252	0.056	1 (sum)
Genotypic mean	76	90	100	115	110	106	101.7 (ave.)
Genotypic deviation	−25.68	−11.68	−1.68	13.32	8.32	4.32	56.0 (var.)
Average excess	$\varepsilon2$: −9.840		$\varepsilon3$: −0.780		$\varepsilon4$: 7.650		
Average effects	−8.945		−0.709		6.955		
Breeding value	−17.891	−9.655	−1.418	−1.991	6.245	13.909	39.7 (var.)
Dominance deviation	−7.789	−2.025	−0.262	15.311	2.075	−9.589	16.4 (var.)

From above, $V_g = 56.0$, $V_e = 554.2 - 56.0 = 498.2$, $V_a = 39.7$, $V_d = 16.4$.
Heritability: broad $= 56.0/554.2 = 0.101$; narrow $= 39.7/554.2 = 0.072$.
$\varepsilon2/\varepsilon4$ is still at greatest risk, and the offspring of $\varepsilon4/\varepsilon4$ are still at greatest risk.

3. a. *AA*: 60 years; *ATh*: 65 years; *ThTh*: 20 years

b. $\mu = (0.9)2(60) + 2(0.9)(0.1)(65) + (0.1)2(20) = 60.5$ years

c. $g_i = G_i - \mu$: *AA*: $60 - 60.5 = -0.5$: *ATh*: 4.5: *ThTh*: −40.5

d. $P_{ij} = \mu + g_i + e_j$; $27 = 60.5 + (-40.5) + e_j$, so $e_j = 7$ years

e. $a_A = (0.9)(-0.5) + (0.1)(4.5) = 0$
$a_{Th} = (0.9)(4.5) + (0.1)(-40.5) = 0$
Because of random mating, average effects = average excesses, so

Breeding value (*AA*) $= a_A + a_A = 0 + 0 = 0$
Breeding value (*ATh*) $= a_A + a_{Th} = 0 + 0 = 0$
Breeding value (*ThTh*) $= a_{Th} + a_{Th} = 0 + 0 = 0$

f. For $p = 0.2$, $\mu = 60$ years, and for $p = 0.05$, $\mu = 60.375$.

	$p = 0.2$			$p = 0.05$		
	AA	*ATh*	*ThTh*	*AA*	*ATh*	*ThTh*
g_i	0	5	−40	−0.375	4.625	−40.375
g_{ai}	2	−3	−8	−0.25	2.25	4.75

Average excess:
$a_A = (0.8)(0) + (0.2)(5) = 1$ $\quad a_A = (0.95)(-0.375) + (0.05)(4.625) = -0.125$
$a_{Th} = (0.8)(5) + (0.2)(-40) = -4$ $\quad a_{Th} = (0.95)(4.625) + (0.05)(-40.375) = 2.375$

The average excess of Th is minimum in absolute value at $p = 0.1$.
The average life span is maximum At $p = 0.1$.

g. At $p = 0.1$, $h^2 = 0$ because all breeding values are zero. At $p = 0.2$:

$$V_a = (0.8)^2(2)^2 + 2(0.8)(0.2)(-3)^2 + (0.2)^2(-8)^2 = 8$$
$$V_g = (0.8)^2(0)^2 + 2(0.8)(0.2)(5)^2 + (0.2)^2(-40)^2 = 72$$
$$V_e = 25, \text{ so } V_p = 72 + 25 = 97$$

Hence, $h^2 = V_a/V_p = 8/97 = 0.0825$.
Similarly, at $p = 0.05$, $h^2 = 0.0190$.
What does $h^2 = 0$ tell you about the inheritance of life span—NOTHING!

CHAPTER 9

Problems

1. Estimate the heritability of male height (in centimeters) from the following data:

Father's Height	Son's Height
180	175
170	172
167	165
177	168
182	180
165	172
172	180
177	172
170	175
179	176

Use the following equation to estimate the correlation coefficient where appropriate:

$$\text{corr}(X, Y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \qquad \bar{x} = \frac{\sum x_i}{n} \qquad \bar{y} = \frac{\sum y_i}{n}$$

2. A population was subjected to selection on a quantitative trait with the following results:

Generation	Total Population Mean	Mean of Selected Parents
0	2.16	2.32
1	2.26	2.34
2	2.30	2.37
3	2.33	2.41
4	2.36	2.47
5	2.40	

a. Estimate the heritability for every generation for which it is possible.

b. Are there any trends in heritability over the generations?

3. Two inbred lines of beans have mean weights of 35 and 65 grams. When crossed, the phenotypic variance of the F_1 was 8, and the phenotypic variance of the F_2 was 41. What is the genetic variance in the F_2 and the broad-sense heritability in the F_2?

4. What is the expected phenotypic covariance between half sibs (offspring who share only one parent in common) in a random-mating population?

5. The highly inbred Canton-S strain of *Drosophila melanogaster* with a high sternopleural bristle number was crossed to the highly inbred Lausanne-S strain with a low bristle number. The F_1's were then used to generate an F_2 and both backcrosses. The mean and variance in sternopleural bristle number in 100 males from each strain or cross are given as follows:

Strain	Mean	Variance
Laussane-S	8.34	0.841
Canton-S	12.12	1.102
F_1	9.48	1.000
F_2	8.89	1.283
BC_1	8.76	1.134
BC_2	10.59	1.180

For the trait of male sternopleural bristle number, use the above data:

a. Estimate the environmental variance.

b. Estimate the additive variance.

c. Estimate the dominance variance.

d. What do the above estimates tell you about the heritability of sternopleural bristle number within the Canton-S and Lausanne-S strains?

6. Following are data on the means and variances of skin color (antilog of reflectance at 685 nanometers) for people of various self-reported backgrounds:

Population	Sample Size	Mean	Variance (×1000)
European	105	0.421	1.09
African	106	0.225	1.05
F_1 African European hybrid	94	0.334	1.59
African backcross	26	0.304	1.71
European backcross	30	0.382	2.00
F_2 hybrid	14	0.346	1.99

For the trait of skin color, use the above data to estimate:

a. The overall dominance deviation, *d*

b. The environmental variance

c. The additive variance

d. The dominance variance

7. Individuals from a population of lizards have a mean adult weight of 40 grams. However, of the lizards that actually bred during a particular year, the mean weight is 50 grams. The mean adult weight of the resulting offspring is 42 grams. Assuming that the parents and offspring experienced identical environmental conditions, what is the heritability of adult weight?

8. The following data were collected from a study on diastolic blood pressure in 612 families. The correlation between a parent and an offspring was 0.183, and the correlation between full sibs was 0.265.

- **a.** Estimate the proportions of the total phenotypic variance that are due to additive effects, dominance effects, and environmental effects.
- **b.** What are the narrow-sense and broad-sense heritabilities of this trait?
- **c.** Repeat parts a and b, but now assume that 0.183 is the correlation between midparent and offspring.

Answers

1. Correlation 0.447607044
heritability = 2 × corelation 0.895214087

2. a.

Generation	Total Population Mean	Mean of Selected Parents	Selection Coefficient	Response	Heritability
0	2.16	2.32	0.16		
1	2.26	2.34	0.08	0.1	0.625
2	2.3	2.37	0.07	0.04	0.5
3	2.33	2.41	0.08	0.03	0.428571429
4	2.36	2.47	0.11	0.03	0.375
5	2.4			0.04	0.363636364

b. The heritability is getting smaller each generation.

3. From the F_1, the environmental variance is 8. Therefore the genetic variance in the F_2 is 41 − 8 = 33, and the broad-sense heritability is 33/41 = 0.805.

4. Half sibs share only a fourth of their alleles. They can share only the half of their alleles they get from their common parent, and Mendelian segregation ensures they share only half of that half. They do not share any genotypic states because that requires that they share both parents. Hence, their expected covariance is $\frac{1}{4}\sigma_a^2$.

5. a. Use equation 9.23 to get $V_e = 0.981$.

b. Use equation 9.24 to get $V_a = 2(1.283) - (1.134 + 1.18) = 0.252$.

c. Use equation 9.25 to get $V_d = 0.05$.

d. Nothing, this experiment only looks at between-strain differences, not within.

6. a. $d = \mu_{F_1} - \frac{1}{2}(\mu_{\text{Euro}} + \mu_{\text{African}}) = 0.334 - 0.323 = 0.011$

b. Use equation 9.23 to get $V_e = 1.23 \times 10^{-3}$.

c. Use equation 9.24 to get $V_a = 0.27 \times 10^{-3}$.

d. Use equation 9.25 to get $V_d = 0.49 \times 10^{-3}$.

7. $S = 50 - 40 = 10$; $\quad R = 42 - 40 = 2$; $\quad h^2 = R/S = \frac{2}{10} = 0.2$

8. For parts a and b:

$$\begin{aligned}
\text{Correlation of parent and offspring } &= (0.5)h^2 = 0.183, \text{ so } h^2 = 0.366.\\
\text{Correlation of sibs } &= (V_a/2 + V_d/4)/V_p = 0.265\\
&= (0.183 + V_d/4)/V_p, \text{ so } V_d/V_p = 0.328.\\
\text{Broad sense } h^2 &= (V_a + V_d)/V_p = 0.366 + 0.328 = 0.694.\\
(V_a + V_d + V_e)/V_p &= 0.694 + V_e/V_p = 1, \text{ so } V_e/V_p = 0.306.
\end{aligned}$$

9. Correlation of midparent and offspring $= \sqrt{\frac{1}{2}}h^2$, so $h^2 = 0.259$; $V_d/V_p = 0.542$; $h^2_{\mathrm{B}} = 0.801$; $V_e/V_p = 0.199$.

CHAPTER 10

Problems

1. Two inbred lines have mean body masses of 50 and 20 kg, respectively. An F_1 is bred and then backcrossed to the larger inbred line. Two marker loci (A and a; B and b, with lowercase alleles coming from the small line), located at the same chromosome with $r = 0.1$, are scored in the backcross progeny with the following results:

Genotype	*AB*/*AB*	*AB*/*Ab*	*AB*/*aB*	*AB*/*ab*
Average size	47	44	40	37

a. What are the expected average sizes for the two recombinant genotypes *AB*/*Ab* and *AB*/*aB* if a QTL for body size exists within this interval and with a recombination rate of 0.03 from the *A* locus?

b. What are the expected average sizes for the two recombinant genotypes *AB*/*Ab* and *AB*/*aB* if a QTL for body size exists within this interval and with a recombination rate of 0.05 from the *A* locus?

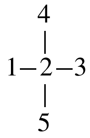
c. Which is the more likely position for the QTL relative to the *A* locus: $r = 0.03$ or $r = 0.05$? Justify your answer.

2. An X-linked locus is surveyed in a sample of 160 males and five haplotypes are discovered. A phenotype is scored for all 160 males, and the mean phenotypes for males bearing haplotype i (along with the estimated variance, s_i^2, and sample size, n_i, for haplotype i) are as follows:

Haplotype	Mean	s^2	n
1	15	95	20
2	20	100	100
3	13	105	20
4	25	90	10
5	18	110	10

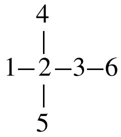
Use the two-sample t test (Appendix 2, equation A2.45) to test the difference between the means of all contrasts between two haplotype categories. Because multiple tests may be used, statistical significance will be corrected by the Bonferroni criteria. Starting with an overall 5% test ($\alpha = 0.05$), no single test will be regarded as significant unless its p value is below 0.05/(number of tests). (Note, ideally the tests corrected in this manner should be independent, which they are not in this case, but this does give a conservative correction and hence is often used.) Are any significant contrasts obtained?

3. Assume the same situation described in problem 2, but now also assume that no recombination has been detected in this X-linked region and that the estimated haplotype tree is



Use the two-sample t test (Appendix 2, equation A2.45) to test the difference between the means of all evolutionarily relevant contrasts of two haplotypes. Because multiple tests may be used, statistical significance will be corrected by the Bonferroni criteria. Starting with an overall 5% test ($\alpha = 0.05$), no single test will be regarded as significant unless its p value is below 0.05/(number of tests). (Note, the evolutionarily relevant contrasts are all asymptotically independent.) Are any significant contrasts obtained?

4. Suppose further sampling of males in the situation described in problem 2 revealed a sixth haplotype with a mean phenotype of 10, estimated variance of 100, and sample size of 5. Suppose further that the haplotype tree is now



The nested-clade design groups haplotypes 1, 2, 4, and 5 together into one-step clade 1–1 and groups haplotypes 3 and 6 together into one-step clade 1-2. The nested analysis first performs all the evolutionarily relevant contrasts among haplotypes *within* each one-step clade and next contrasts the two one-step clades against one another. Use t tests to perform all the necessary contrasts required by the nested design and do a Bonferroni correction (all contrasts in a nested-clade analysis are asymptotically independent, so a simple Bonferroni correction is appropriate). To perform the contrast of the one-step clades, you will need to pool data from two or more haplotype categories. Let J be a set of haplotypes found in one-step clade $1 - j$. Then the pooled mean phenotype for clade $1 - j$ is

$$\text{Mean}(j) = \frac{\sum_{i \text{ in } J} n_i \bar{x}_i}{\sum_{i \text{ in } J} n_i}$$

The pooled estimated variance is

$$s^2(j) = \frac{\sum_{i \text{ in } J} n_i s_i^2}{\sum_{i \text{ in } J} n_i}$$

and the pooled sample size is $\sum_{i \text{ in } J} n_i$. If any significant contrasts are detected, identify the branch(es) in the haplotype tree with the significant phenotypic association(s).

5. Perform one round of tree scanning on the problem described in problem 4. In this case, the Bonferroni correction will be overly conservative because the contrasts are not independent, but use this correction anyway. If any significant contrasts are detected, identify the branch(es) in the haplotype tree with the significant phenotypic association(s).

Answers

1. For parts a and b, use Table 10.1 (note, you have to reverse the genotypes because this is the opposite backcross to the one shown in that table):

Genotype	*AB*/*AB*	*AB*/*Ab*	*AB*/*aB*	*AB*/*ab*	*r* from *A* to QTL
Average size	47	44	40	37	
Expected value given *r*		44	40		0.03
		42	42		0.05

c. $r = 0.03$ is the more likely location as it explains the recombinant means better.

2. There are 10 possible contrasts:

Contrast	*t*	df	*p*
1 vs. 2	2.0495	118	0.042629141
1 vs. 3	0.1432	38	0.886870876
1 vs. 4	−0.8449	28	0.405336846
1 vs. 5	−0.2452	28	0.808114383
2 vs. 3	2.8463	118	0.005217409
2 vs. 4	−1.5139	108	0.132977928
2 vs. 5	0.6005	108	0.54941408
3 vs. 4	−0.9789	28	0.336001289
3 vs. 5	−0.3954	28	0.695547805
4 vs. 5	0.5078	18	0.617736579
	Bonferroni for 10 tests:		0.005

None of the tests have p values below the Bonferroni p level.

3. Here are all the evolutionarily relevant contrasts and test results:

Contrast	*t*	df	*p*
1 vs. 2	2.0495	118	0.042629141
3 vs. 2	2.8463	118	0.005217409
4 vs. 2	−1.5139	108	0.132977928
5 vs. 2	0.6005	108	0.54941408
	Bonferroni for 4 tests:	0.0125	

The only test that satisfies the Bonferroni cutoff is the contrast between haplotypes 3 and 2.

4. First are the evolutionarily relevant haplotype contrasts within clades 1–1 and 1–2:

		t	df	p
Clade 1–1	1 vs. 2	2.0495	118	0.042629141
	4 vs. 2	−1.5139	108	0.132977928
	5 vs. 2	0.6005	108	0.54941408
Clade 1–2	3 vs. 6	0.5880	23	0.562275267

To contrast the one-step clades, we must first pool:

	Mean	S^2	N
Clade 1–1	19.5	99.28571429	140
Clade 1–2	12.4	104	25

Finally, we do the final t test to contrast 1–1 vs. 1–2:

	t	df	p
1–1 vs. 1–2	3.270349555	163	0.001310713

There are a total of five t tests in the nested-clade analysis, so the Bonferroni 5% p value is 0.01. The only contrast that is significant is the contrast between clades 1–1 and 1–2. Hence, the only branch in the tree associated with a significant phenotypic change is the branch interconnecting haplotypes 2 and 3.

5. To perform a tree scan, we first must pool some haplotypes:

Pooled Haplotypes	Mean	s^2	n
2, 3, 4, 5, 6	18.89655172	100.6896552	145
1, 2, 3, 5, 6	18	100.6451613	155
1, 2, 3, 4, 6	18.4516129	99.35483871	155
1, 2, 4, 5	19.5	99.28571429	140
3, 6	12.4	104	25
1, 2, 3, 4, 5	18.6875	100	160

Next, do all the contrasts made by cutting one branch and test:

Contrast	t	df	p
1 vs. 2, 3, 4, 5, 6	−1.633353013	163	0.104324891
4 vs.1, 2, 3, 5, 6	2.144851991	163	0.033445955
5 vs.1, 2, 3, 4, 6	−0.138456988	163	0.890050165
1, 2, 4, 5 vs. 3, 6	3.270349555	163	0.001310713
1, 2, 3, 4, 5 vs. 6	−1.912924542	163	0.057510967
	Bonferroni for 5 tests:	0.01	

Only one test is significant, and it localizes the association to the branch interconnecting haplotypes 2 and 3.

CHAPTER 11

Problems

1. A population is polymorphic at an autosomal locus with two alleles, *A* and *a*. The probability is 0.1 for an *AA* zygote surviving to adulthood, 0.08 for an *Aa* zygote, and 0.12 for an *aa* zygote. The probability is 0.8 for an *AA* adult to successfully mate, 0.9 for an *Aa* adult, and 0.7 for an *aa* adult. A mated *AA* individual produces an average of 10 offspring, a mated *Aa* individual produces 9 offspring, and a mated *aa* individual produces 8 offspring.
   - **a.** What are the expected numbers of offspring for *AA*, *Aa*, and *aa* zygotes?
   - **b.** Assign a fitness of 1 to *aa*. What are the relative fitnesses of the *AA* and *Aa* genotypes?

2. The zygote-to-adult relative viabilities are 1, 0.5, and 0.25 for the *AA*, *Aa*, and *aa* genotypes, respectively. Assume random mating.
   - **a.** Starting with the zygotic frequency of *A* being 0.2, what are the genotype frequencies in the adult population?
   - **b.** Suppose you did not know the fitnesses but did know the population was randomly mating. Suppose further that you can *only* observe the adult population. Some workers estimate relative viabilities as deviations from Hardy–Weinberg genotype frequencies. Given only what you can observe about this population, what are your estimates of relative viabilities using the Hardy–Weinberg deviation technique?

3. *Drosophila pseudoobscura* is polymorphic for an autosomal inversion, with the two states designated *ST* and *CH*. These two rearrangements behave like alleles in a single-locus system with respect to Mendelian inheritance. A large number of *ST*/*CH* flies were put in a cage, and the next generation consisted of 69 *CH*/*CH*, 234 *ST*/*CH*, and 97 *ST*/*ST* flies.
   - **a.** What are the relative fitnesses of the three karyotypes when the fitness of *ST*/*CH* is set to 1?
   - **b.** What is the evolutionary fate of this experimental population under random mating in subsequent generations assuming that selection is the only evolutionary force?

4. The peppered moth is polymorphic at an autosomal locus determining a melanic form. Let the three genotypes be *MM*, *Mm*, and *mm* with estimated viabilities of 0.9, 0.95, and 1, respectively. Assume the frequency of the *m* allele was 0.1 in Oxford in 1940.
   - **a.** What do you expect the frequency of *m* to be in 1941 (one generation per year) under the assumption of random mating?
   - **b.** What do you expect the frequency of *m* to be in 1941 under the assumption of assortative mating yielding an f for this locus of 0.2?

5. The following fitnesses were estimated for humans living in a malarial environment in West Africa as a function of their genotype at the hemoglobin β-chain locus:

Genotype	*AA*	*AS*	*SS*	*AC*	*SC*	*CC*
Fitness	0.9	1	0.2	0.9	0.71	1.31

Given these fitnesses and the assumption of random mating in an infinitely large population, calculate each population's average fitness and the average excess and rate of change of allele frequency for all three alleles for populations with the following initial gene pools:

	ALLELE FREQUENCY		
	A	*S*	*C*
a.	0.9	0.05	0.05
b.	0.85	0.05	0.10
c.	0.75	0.05	0.20

d. Do the initial allele frequencies affect the course of adaptive evolution in the above cases (that is, are the signs in the changes of allele frequencies the same in parts a, b, and c)?

e. Redo parts a, b, c, and d keeping all the same fitnesses and gene pools, but now assuming an inbreeding system of mating for which $f = 0.05$ (measured as a deviation from Hardy–Weinberg genotype frequencies).

f. Does this change in population structure alter the course of adaptive evolution in any of the cases (once again, in terms of the sign of allele frequency change)?

6. Two populations were studied in two environments with the following results (note, treat the two populations as independent evolutionary units):

	Environment 1			Environment 2		
Genotype	*AA*	*AB*	*BB*	*AA*	*AB*	*BB*
Zygote number	160	480	360	160	480	360
Adult number	90	450	300	150	400	340

a. Assuming that viability is the only component of fitness affected by genetic variability at this locus and that generations are discrete, estimate the relative fitnesses of the three genotypes in each environment using the convention that the fitness of the *AB* heterozygote is 1.

b. Assuming the population structure remains constant and that drift can be ignored, what is the population structure and what is the initial response to selection under the two environments (i.e., which alleles are favored and what is Δp)?

c. Does a polymorphic equilibrium point exist for each of the environments (that is, $\Delta p = 0$ for $0 < p < 1$)? If so, do the initial changes in allele frequency calculated above move the population away or toward any polymorphic equilibria that may exist?

7. A trait x is related to fitness by the function $w(x) = 3x - x^2$ for $0 \leq x \leq 3$.

a. What value of x results in maximum fitness for bearers of that phenotypic value (hint, remember your calculus and use first and second derivatives)?

b. Consider a population at selective equilibrium with a positive phenotypic variance. Do you expect this population at selective equilibrium to have the average trait value of x that optimizes $w(x)$ that you calculated above? Why or why not?

8. A population has an average fitness of 0.9, a variance of fitness of 3, and a heritability of fitness of 0.1.

a. What is the average fitness in the next generation?

b. What will be the heritability of fitness at equilibrium?

9. Consider the following genotypes and fitnesses:

AA	*Aa*	*aa*
1	0.4	1.2

a. Assuming a random-mating population, does a nontrivial (i.e., $p \neq 0$ or $p \neq 1$) equilibrium exist? (Hint: The average excesses of A and a are equal at equilibrium.) If such a nontrivial equilibrium does exist, show whether or not it is stable (a numerical demonstration is fine, using $p_{\text{eq}} \pm 0.1$).

b. Redo part a, but now assume a system of mating with $f = 0.4$.

10. A trait x is related to fitness by the function $w(x) = 10x - x^2$ for $0 < x < 10$. What value of x results in maximum fitness for bearers of that phenotypic value? Consider a population at selective equilibrium. At this selective equilibrium, the phenotypic variance of x is 0. What is the mean of x at selective equilibrium, and what is the average fitness of the population?

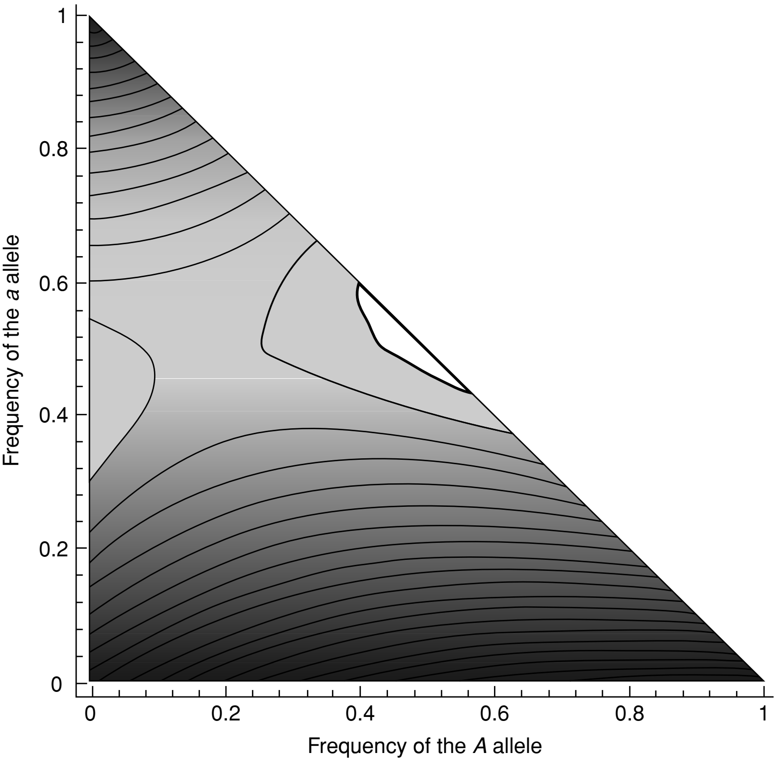
11. The following fitnesses were estimated for a locus with three alleles (A,a,α) in a population:

Genotype	*AA*	*Aa*	*aa*	*Aα*	*aα*	*αα*
Fitness	0.2	1.7	0.3	0.1	1.4	0.6

Given these fitness and the assumption of random mating in an infinitely large population, calculate the rate of change of average fitness using Fisher's fundamental theorem of natural selection for the following initial gene pools:

	ALLELE FREQUENCY		
	A	*a*	*α*
a.	0.55	0.40	0.05
b.	0.35	0.60	0.05
c.	0.05	0.20	0.75

Below is a contour plot of average fitness versus the allele frequencies of A and a ($p + q \leq 1$), with darker shading corresponding to lower average fitness. How many local fitness peaks are in this adaptive landscape, and do they represent stable polymorphic equilibria in a local sense? Give the approximate gene pool states at these equilibria as extrapolated from the plot below. Toward which fitness peak does natural selection take the population in parts a, b, and c. What do the average excesses in parts a, b, and c say about the direction of the initial response to selection?



12. For the following fitnesses and genotype frequency arrays, calculate the change in the frequency of the *A* allele in the following generation:

	Genotype Frequency			Fitness		
	AA	*Aa*	*Aa*	*AA*	*Aa*	*aa*
a.	0.4	0.3	0.3	100	135	140
b.	0.8	0.05	0.15	0.005	0.003	0.003
c.	0.35	0.35	0.3	π	2π	6π

13. Consider the following pleiotropic traits influenced by a single autosomal locus with two alleles that is found in a random–mating population:

Genotype	*AA*	*Aa*	*aa*
Viability	0.9	0.5	0.2
Mating success	1	1	1
Fecundity	1	3	4

Calculate the equilibrium allele frequency. Calculate the additive variance in fitness (do not use relative fitness), viability, and fecundity at equilibrium. Are these additive variances consistent with Fisher's fundamental theorem?

Answers

1.

	AA	Aa	aa
l_x	0.1	0.08	0.12
c	0.8	0.9	0.7
b	18	20	10
a. w	1.44	1.44	0.84
b. Relative fitness	1.714285714	1.714285714	1.00

2.

	AA	Aa	aa	p
Zygotic Hardy–Weinberg genotype frequency	0.04	0.32	0.64	0.2
Fitnesses (average fitness = 0.36)	1	0.5	0.25	
a. Adult genotype frequency:	0.1111	0.4444	0.4444	0.3333
Expected Hardy–Weinberg given adult p	0.1111	0.4444	0.4444	

b. Since the adult genotype frequencies are equal to the expected Hardy–Weinberg frequencies given the adult allele frequency, the estimated relative viabilities would all be equal to 1.

3.

	CH/CH	ST/CH	ST/ST	Sum
Observed	69	234	97	400
Expected	100	200	100	
Fitness = observed/expected	0.69	1.17	0.97	
a. Relative fitness	0.58974359	1	0.829059829	
Selection coefficient	0.41025641		0.170940171	

b. Using equation 11.13, a balanced polymorphism, with p at equilibrium = 0.294117647.

4.

	MM	Mm	mm
Hardy–Weinberg genotype frequency	0.81	0.18	0.01
Assortative mating genotype frequency	0.828	0.144	0.028
Fitness	0.9	0.95	1

Average fitness ($f = 0$): 0.91
Average fitness ($f = 0.2$): 0.91

		$f = 0$	$f = 0.2$
Average excess:	A	−0.005	−0.006
	a	0.045	0.054

Δq for $f = 0$ 0.004945055 So $q' = 0.104945055$
Δq for $f = 0.2$ 0.005934066 So $q' = 0.105934066$

5. Under Random Mating

a.

Genotype	*AA*	*AS*	*SS*	*AC*	*SC*	*CC*	
Fitness	0.9	1	0.2	0.9	0.71	1.31	$\bar{W} = 0.907325$
Frequency	0.81	0.09	0.0025	0.09	0.005	0.0025	

Allele	*A*	*S*	*C*
Average excess	−0.002325	0.038175	0.003675
$\Delta p = pa_x/\bar{W}$	−0.0023062	0.00210371	0.00020252

b.

Genotype	*AA*	*AS*	*SS*	*AC*	*SC*	*CC*	
Fitness	0.9	1	0.2	0.9	0.71	1.31	$\bar{W} = 0.90895$
Frequency	0.7225	0.085	0.0025	0.17	0.01	0.01	

Allele	*A*	*S*	*C*
Average excess:	−0.00395	0.02205	0.02255
$\Delta p = pa_x/\bar{W}$	−0.0036938	0.00121294	0.00248088

c.

Genotype	*AA*	*AS*	*SS*	*AC*	*SC*	*CC*	
Fitness	0.9	1	0.2	0.9	0.71	1.31	$\bar{W} = 0.91835$
Frequency	0.5625	0.075	0.0025	0.3	0.02	0.04	

Allele	*A*	*S*	*C*
Average excess:	−0.01335	−0.01635	0.05415
$\Delta p = pa_x/\bar{W}$	−0.0109027	−0.0008902	0.01179289

d. The initial array of allele frequencies does affect the course of adaptive evolution, with the sign in the change of the *S* allele frequency being reversed in part c from Parts a and b.

e. Under Inbreeding with $f = 0.05$
Part a:

Genotype	*AA*	*AS*	*SS*	*AC*	*SC*	*CC*	
Fitness	0.9	1	0.2	0.9	0.71	1.31	$\bar{W} = 0.906234$
Frequency	0.8145	0.0855	0.004875	0.0855	0.00475	0.004875	

Allele	*A*	*S*	*C*
Average excess	−0.0014837	0.00199125	0.02471625
$\Delta p = pa_x/\bar{W}$	−0.0014735	0.00010986	0.00136368

Part b:

Genotype	*AA*	*AS*	*SS*	*AC*	*SC*	*CC*	
Fitness	0.9	1	0.2	0.9	0.71	1.31	$\bar{W} = 0.9088025$
Frequency	0.729	0.081	0.0049	0.1615	0.0095	0.0145	

Allele	*A*	*S*	*C*
Average excess	−0.0040524	−0.0143525	0.0416225
$\Delta p = pa_x/\bar{W}$	−0.0037903	−0.0007896	0.00457993

Part c:

Genotype	*AA*	*AS*	*SS*	*AC*	*SC*	*CC*	
Fitness	0.9	1	0.2	0.9	0.71	1.31	$\bar{W} = 0.9197825$
Frequency	0.5147	0.07125	0.001	0.2565	0.0135	0.06288	

Allele	*A*	*S*	*C*
Average excess	−0.0150325	−0.0528825	0.0695925
$\Delta p = pa_x/\bar{W}$	−0.0122577	−0.0028747	0.01513238

Part d: The initial array of allele frequencies does affect the course of adaptive evolution, with the sign in the change of the *S* allele frequencies being reversed in part a from parts b and c.

Part f: The course of evolution is altered by the system of mating, with the sign of the change in the *S* allele frequency changing in part b from + (random mating) to − (inbreeding).

6.

	Environment 1			Environment 2		
Genotype	*AA*	*AB*	*BB*	*AA*	*AB*	*BB*
Zygote no. (Z#)	160	480	360	160	480	360
Adult no. (A#)	90	450	300	150	400	340
Viability = A#/Z#	0.5625	0.9375	0.8333	0.9375	0.8333	0.9444
a. Relative fitness	0.6	1	0.8889	1.125	1	1.133

b. Both populations are in Hardy–Weinberg and have $p_A = 0.4$, with genotype frequencies of 0.16, 0.48, and 0.36, respectively.

	Environment 1	Environment 2
$\bar{W}$	0.896	1.0678

Allele	*A*	*B*	*A*	*B*
Average excess	−0.056	0.0373	−0.0178	0.0112
$\Delta p = pa_x/\bar{W}$	−0.025	0.025	−0.00637	0.00637

c. Polymorphic equilibrium is defined by $a_A = a_B$ or $pw_{AA} + (1-p)w_{AB} = pw_{AB} + (1-p)w_{BB}$. Solving the above equation for p given the fitnesses,

For environment 1, $p_{\text{eq}} = 0.2174$ (or for the *B* allele, 0.7826).
For environment 2, $p_{\text{eq}} = 0.5161$ (or the for *B* allele, 0.4839).

In environment 1, $p_A = 0.4$ is higher than the equilibrium, and the negative average excess causes selection to get closer to the equilibrium.

In environment 2, $p_A = 0.4$ is lower than the equilibrium, and the negative average excess causes selection to get farther away from the equilibrium.

7. a. To find the maximum of $w(x)$, take the first derivative and set it to 0: $w'(x) = 3 - 2x = 0$, so $x_{\text{max}} = \frac{3}{2} = 1.5$.

To confirm this is a maximum, take the second derivative, $w''(x) = -2$. This is negative, so it is a maximum.

b. For a population at selective equilibrium, you do not expect $\bar{x} = x_{\text{max}}$ because $w(x)$ is nonlinear.

8. a. From FFTNS, $\Delta\bar{W} =$ (additive variance of fitness)$/\bar{W} = h^2$ (variance of fitness)$/\bar{W} = 0.1(3)/0.9 = 0.333$. Hence,

$$\bar{W}' = \bar{W} + \Delta\bar{W} = 0.9 + 0.333 = 1.233$$

b. At equilibrium, $h^2 = 0$ (from FFTNS).

9. a. $a_A = p(1) + (1-p)(0.4) - \bar{W} = p(0.4) + (1-p)(1.2) - \bar{W} = a_a$ at equilibrium, so

$$\begin{aligned} p(1) + (1-p)(0.4) = p(0.4) + (1-p)(1.2) &\Rightarrow p - 0.4p - 0.4p + 1.2p \\ &= -0.4 + 1.2,\ 1.4p = 0.8 \Rightarrow p_{\text{eq}} = 0.8/(1.4) = 0.571 \end{aligned}$$

at $p = 0.671, a_A = 0.054;$ at$p = 0.471 a_A = -0.061 \Rightarrow$ unstable.

b. With $f = 0.4$:

$$\begin{aligned} a_A &= [f + p(1-f)](1) + (1-p)(1-f)(0.4) - \bar{W} \\ &= p(1-f)(0.4) + [f + (1-p)(1-f)](1.2) - \bar{W} = a_a \end{aligned}$$

so

$$\begin{aligned} &[0.4 + p(0.6)](1) + (1-p)(0.6)(0.4) = p(0.6)(0.4) \\ &\quad + [0.4 + (1-p)(0.6)](1.2) \Rightarrow 0.84p = 0.560 \Rightarrow p_{\text{eq}} = 0.667 \end{aligned}$$

at $p = 0.767, a_A = 0.026$; at $p = 0.567, a_A = -0.036 \Rightarrow$ unstable.

10. $w(x)' = 10 - 2x$. To find max/min, set the derivative to 0: $10 - 2x = 0 \Rightarrow x = 5$. To see if max or min (or saddle point), take second derivative to get $-2 \Rightarrow$ max.

Although $w(x)$ is nonlinear, there is no phenotypic variance, so from equation 11.26, $x_{\text{eq}} = 5 = x_{\text{opt}}$ and $\bar{W} = w(5) = 25$.

11. a.

Genotype	AA	Aa	aa	Ab	$a\alpha$	$\alpha\alpha$	Sum or Average
Fitness	0.2	1.7	0.3	0.1	1.4	0.6	0.9195
Genotype frequency	0.3025	0.44	0.16	0.055	0.04	0.0025	1
Allele	A	a	α				
Allele frequency	0.55	0.4	0.05				1

	A	a	α
Average excess	−0.1245	0.2055	−0.2745

	AA	*Aa*	*aa*	*Aα*	*aα*	*αα*	Additive variance:
Breeding value (of fitness)	−0.249	0.081	0.411	−0.399	−0.069	−0.549	0.0583695

FFTNS rate of change of average fitness: 0.063479608

b. Genotype	*AA*	*Aa*	*aa*	*Aα*	*aα*	*αα*	
Fitness	0.2	1.7	0.3	0.1	1.4	0.6	0.9355
Genotype frequency	0.1225	0.42	0.36	0.035	0.06	0.0025	1

Allele	*A*	*a*	*α*	
Allele frequency	0.35	0.6	0.05	1

	A	*a*	*α*
Average excess	0.1595	−0.0905	−0.0305

	AA	*Aa*	*aa*	*Aα*	*aα*	*αα*	Additive variance:
Breeding value	0.319	0.069	−0.181	0.129	−0.121	−0.061	0.0277295

FFTNS rate of change of average fitness: 0.029641368

c. Genotype	*AA*	*Aa*	*aa*	*Aα*	*aα*	*αα*	
Fitness	0.2	1.7	0.3	0.1	1.4	0.6	0.8115
Genotype frequency	0.0025	0.02	0.04	0.075	0.3	0.5625	1

Allele	*A*	*a*	*α*	
Allele frequency	0.05	0.2	0.75	1

	A	*a*	*α*
Average excess	−0.3865	0.3835	−0.0765

	AA	*Aa*	*aa*	*Aα*	*aα*	*αα*	Additive variance:
Breeding value	−0.773	−0.003	0.767	−0.463	0.307	−0.153	0.0825455

FFTNS rate of change of average fitness: 0.101719655

In the contour plot, there are two peaks separated by a saddle. They are locally stable because they are peaks (all approaches to them represent increasing $\bar{W}$). They are approximately at p_A of about 0.5 and p_a of about 0.5 and at p_A of 0 and p_a of about 0.45.

In Part a you go to the taller peak (about 0.5, 0.5). Note, you start out with the frequency of *A* too high for this peak, and its average excess is negative (which takes you closer), and *a* is too low but its average excess is positive (which takes you closer).

In Part b you also go to the taller peak. In this case you start out with the frequency of *A* too low for this peak, and the average excess is positive; the frequency of *a* is too high but its average excess is negative.

In Part c you go to the lower peak (0, 0.45). You start out with the frequency of A too high for this peak, and its average excess is negative (which takes you closer), and the frequency of a is too low but its average excess is positive (which takes you closer).

12. **a.** $p = 0.55, \bar{W} = 122.5, a_A = -12.95, \Delta p = pa_A/\bar{W} = -0.058$
b. $p = 0.825, \bar{W} = 0.0535, a_A = -0.0041, \Delta p = -0.063$
c. $p = 0.525, \bar{W} = 2.85\pi, a_A = -1.517\pi, \Delta p = -0.279$

13. First, calculate the fitnesses by multiplying the fitness components:

Genotype	*AA*	*Aa*	*aa*
Viability	0.9	0.5	0.2
Mating success	1	1	1
Fecundity	1	3	4
Fitness	0.9	1.5	0.8

In a random-mating population at equilibrium, $a_A = a_a$, so

$$p(0.9) + (1 - p)(1.5) - \bar{W} = p(1.5) + (1 - p)(0.8) - \bar{W}$$
$$-1.3p = -0.7$$
$$p_{\text{eq}} = 0.5385$$

At equilibrium, $\bar{W} = 1.1769$, $a_A = a_a = 0$, so all breeding values (additive genotypic deviations) are zero, so the additive variance in fitness is zero, consistent with the fundamental theorem.

At equilibrium, average viability = 0.5521, $a_A = 0.1633$, $a_a = -0.1905$, so

Genotype	*AA*	*Aa*	*aa*
Breeding value	0.3266	−0.0272	−0.3811

So the additive variance (variance of breeding values) is 0.0622. This result is also consistent with the fundamental theorem because viability is not the same as fitness and the FFTNS only ensures that the additive variance of fitness is zero at equilibrium.

At equilibrium, the average fecundity is 2.6331, $a_A = -0.7101$, $a_a = 0.8284$, so

Genotype	*AA*	*Aa*	*aa*
Breeding value	−1.4201	0.1183	1.6568

So the additive variance (variance of breeding values) is 1.1764. This is also consistent with FFTNS because fecundity is not the same as fitness.

CHAPTER 12

Problems

1. Assume $h = 0.03$ for a mutation that is lethal when homozygous and $h = 0.4$ for a mildly deleterious mutation when homozygous ($s = 0.05$). In both cases, the mutation rate to the deleterious form is 10^{-5}.

- **a.** Assuming random mating, what are the equilibrium frequencies for the mutant allele in these two cases?
- **b.** Assuming $f = 0.1$ and that both mutants are now completely recessive, what are the equilibrium frequencies for the mutant allele in these two cases?

2. Cystic fibrosis (CF) is a fatal disease caused by a recessive mutation. Individuals homozygous for the *CF* allele die before reproducing. The *CF* allele has a frequency of 0.02 in a randomly mating human population. Assume that this is an equilibrium frequency.

- **a.** Assume in addition that the *CF* heterozygotes have the same fitness as the normal-allele homozygotes. Estimate the mutation rate from the normal to *CF* alleles.
- **b.** Assume instead that the *CF* heterozygotes have greater viability than either homozygote. Assume further that selection is sufficiently strong that the effects of mutation on the equilibrium frequency can be ignored. Estimate the selection coefficient of the normal-allele homozygote.
- **c.** There has been controversy over whether the *CF* frequency reflects a mutation/selection balance or an overdominance (heterozygote superiority) balance. In light of your answers to part a and b and your knowledge of mutation rates in general, which seems more reasonable under the assumption of equilibrium?

3. A population of snakes lives on an island and has the following fitnesses at a locus with two alleles, *A* and *a*:

Genotype	*AA*	*Aa*	*aa*
Fitness	1	0.8	0.2

This island population also receives some gene flow from the mainland population, with 10% of the gametes coming in from the mainland every generation. The frequency of the *a* allele is maintained at 0.7 by overdominant selection in the mainland environment.

- **a.** Assuming random mating and a frequency of *a* of 0.2 on the island, what is the frequency of the *a* allele on the island in the next generation?
- **b.** Same as part a, but now assume the frequency of *a* on the island is 0.3.
- **c.** Assuming avoidance of inbreeding with $f = -0.2$ and a frequency of *a* of 0.2 on the island, what is the frequency of the *a* allele on the island in the next generation?
- **d.** Same as part c, but now assume the frequency of *a* on the island is 0.3.
- **e.** Under which system of mating does the equilibrium between selection and gene flow yield an allele frequency between 0.2 and 0.3? Justify your answer.

4. A codominant advantageous mutation, *a*, occurs in a random-mating population with the following fitnesses: *AA*, 1; *Aa,* 1.05, and *aa*, 1.1. What is the probability of fixation of this mutation when:

- **a.** The population is ideal (all effective sizes = census size) and of size 50?
- **b.** The population is ideal (all effective sizes = census size) and of size 500?
- **c.** The variance effective size is 50 and the census size is 200?

5. A molecular genetic survey is conducted at a homologous protein-coding locus in two closely related species, and the distribution of all mutations is mapped onto the branches of the resulting haplotype tree with the following results:

	Silent Substitutions	Replacement Substitutions
Tip	20	7
Interior	15	16
Fixed	18	24

Use a contingency chi square (Appendix 2) to test the hypothesis of neutrality. If you reject that hypothesis at or below the 5% level of significance, describe the type of selection that is occurring at this locus.

Now test the null hypothesis of neutrality with these data using the contingency test of McDonald and Kreitman. Contrast the results to the previous contingency test.

6. Consider a diploid population with two alleles (A and a) at an autosomal locus and an initial frequency of the A allele of 0.8. Now consider the following genotypes and fitnesses:

AA	Aa	aa
1	0.5	1.5

There are two stable equilibria under selection in this population, at $p = 0$ and $p = 1$.

a. Suppose the population is effectively infinite in size and randomly mating but receives a portion m of its genes from another population fixed for allele a. What is the minimum value of m that will cause the population in the next generation to be in the domain of the $p = 0$ equilibrium?

b. Same as part a, but now assume the system of mating has $f = -0.1$.

Answers

1. a. Use equation 12.8 to get $q_{\text{eq}} = 10^{-5}/[(0.03)(1)] = 0.00033$ for the first case and $q_{\text{eq}} = 10^{-5}/[(0.4)(0.05)] = 0.0005$ for the second case.

b. Use equation 12.10 to get $q_{\text{eq}} = 10^{-5}/[(0.1)(1)] = 0.0001$ for the first case and $q_{\text{eq}} = 10^{-5}/[(0.1)(0.05)] = 0.002$ for the second case.

2. a. Using equation 12.7 and noting that $s = 1$, the estimate of the mutation rate at equilibrium is $\mu = q^2 = 0.0004$.

b. Using equation 11.13 and noting that $s = 1$, the estimate of t is $0.02/0.98 = 0.0204$.

c. 0.0004 would represent an extremely high mutation rate and therefore is unlikely. A modest selection coefficient of only 2% could explain the results under the overdominance hypothesis, so this explanation seems more reasonable.

3. a. Use equation 12.13 with $p_2 = 0.7$ and $m = 0.1$. Under random mating with $p_1 = 0.2$, the average fitness is 0.904, the average excess of the a allele is −0.224, so from equation 12.13, the change in p is 0.00044248 and the frequency in the next generation is 0.20044248.

b. Use equation 12.13 with $p_2 = 0.7$ and $m = 0.1$. Under random mating with $p_1 = 0.3$, the average fitness is 0.844, the average excess of the a allele is −0.224, so from equation 12.13, the change in p is -0.0396209 and the frequency in the next generation is 0.26037915.

c. Use equation 12.13 with $p_2 = 0.7$ and $m = 0.1$. With $f = -0.2$ and $p_1 = 0.2$, the average fitness is 0.9168, the average excess of the a allele is -0.1408, so from equation 12.13, the change in p is 0.01928447 and the frequency in the next generation is 0.21928447.

d. Use equation 12.13 with $p_2 = 0.7$ and $m = 0.1$. With $f = -0.2$ and $p_1 = 0.3$, the average fitness is 0.8608, the average excess of the a allele is -0.1568, so from equation 12.13, the change in p is -0.0145468 and the frequency in the next generation is 0.2853516.

e. Under both systems of mating, because at $p = 0.2$ the allele frequency increases but at $p = 0.3$ the allele frequency decreases.

4. a. Note that $s = 0.05$; then use equation 12.16 to obtain $u = 0.0951669$.

b. $u = 0.09516258$

c. Use equation 12.15 to obtain $u = 0.02469$.

5. Get the marginals:

	Silent Substitutions	Replacement Substitutions	Sum
Tip	20	7	27
Interior	15	16	31
Fixed	18	24	42
Sum	53	47	100

Then generate the expected values:

	Silent Substitutions	Replacement Substitutions
Tip	14.31	12.69
Interior	16.43	14.57
Fixed	22.26	19.74

Calculate the chi square from the observed and expected values as 6.81 with two degrees of freedom, which is significant at the 5% level ($p = 0.033$). Therefore, we reject the null hypothesis of neutrality. The silent substitutions are disproportionately on the tips and the replacements on the fixed branches, so this implies positive selection favoring amino acid change in this protein-coding locus.

The McDonald–Kreitman test pools the categories of tip and interior into polymorphic, so this results in a 2 × 2 contingency table:

	Silent	Replacement	Sum
Polymorphic	35	23	58
fixed	18	24	42
Sum	53	47	100

The expecteds are

	Silent	Replacement
Polymorphic	30.74	27.26
Fixed	22.26	19.74

yielding a chi square of 2.99 with one degree of freedom, which is not significant at the 5% level ($p = 0.0837$). In contrast to the previous test, here the null hypothesis of neutrality cannot be rejected.

6. a. $a_A = p(1) + (1-p)(0.5) - \bar{W} = p(0.5) + (1-p)(1.5) - \bar{W} = a_a$ at the unstable equilibrium, so $p + (1-p)(0.5) = p(0.5) + (1-p)(1.5) \Rightarrow p - 0.5p - 0.5p + 1.5p = -0.5 + 1.5 \Rightarrow 1.5p = 1 \Rightarrow p_{\text{eq}} = 1/1.5 = \frac{2}{3}$.

At $p = 0.8$, $\bar{W} = 0.86$ and $a_A = 0.056$.

$\Delta p = pa_A/\bar{W} + \Delta p$ (gene flow) $= pa_A/\bar{W} - m(p_1 - p_2) = 0.037 - m(0.8 - 0) = 0.037 - 0.8m$.

To put p into the domain of $p_{\text{eq}} = 0$, $p' < \frac{2}{3}$, or $\Delta p < 0.667 - 0.8 = -0.133$; so, $0.037 - 0.8m < -0.133 \Rightarrow -0.8m < -0.170 \Rightarrow m > 0.213$.

b. $a_A = [f + p(1-f)](1) + (1-p)(1-f)(0.5) - \bar{W} = p(1-f)(0.5) + [f + (1-p)(1-f)](1.5) - \bar{W} = a_a$, so $[-0.1 + p(1.1)] + (1-p)(1.1)(0.5) = p(1.1)(0.5) + [-0.1 + (1-p)(1.1)](1.5) \Rightarrow 1.65p = 1.05 \Rightarrow p_{\text{eq}} = 0.636$.

At $p = 0.8$, $\bar{W} = 0.836$ and $a_A = 0.076$.

$\Delta p = pa_A/\bar{W} + \Delta p$ (gene flow) $= pa_A/\bar{W} - m(p_1 - p_2) = 0.052 - m(0.8 - 0) = 0.052 - 0.8m$.

To put p into the domain of $p_{\text{eq}} = 0$, $p' < 0.636$, or $\Delta p < 0.636 - 0.8 = -0.164$, so $0.052 - 0.8m < -0.164 \Rightarrow -0.8m < -0.216 \Rightarrow m > 0.27$.

CHAPTER 13

Problems

1. An allele A shows meiotic drive over allele a such that in heterozygotes there is a 4 : 1 segregation in favor of A. Assume for now that this is the only phenotype associated with A.

a. Starting with an allele frequency of 0.05 for A in a random-mating population, what is the frequency of A in the next generation?

b. Same as part a, but now assume nonrandom mating with $f = 0.5$.

c. Assume random mating again but also assume that AA homozygotes are lethal but Aa and aa genotypes are equally fit. What is the frequency of A in the next generation?

2. A tandem, multigene family with 100 copies has a per-locus neutral mutation rate of 10^{-5} per generation in an ideal population of size 10,000. The probability of one copy converting a paralogous copy to its state is 10^{-4} per generation.

a. What is the rate of neutral evolution for the entire multigene family?

b. What is the expected time to fixation (in all individuals in the population and at all paralogous sites) of a new neutral mutation?

c. Which is the limiting step in the total fixation time given above: orthologous fixation or paralogous fixation?

3. Given the following fitnesses, show whether or not the polymorphism is protected. If the polymorphism is not protected, identify the alleles that are not protected. In all cases, assume random mating.

a. Competitive interactions:

Genotype	*AA*	*Aa*	*aa*
AA	0.95	0.9	1.2
Aa	0.9	1	1.7
aa	1.2	1.7	1.6

b. Fertility:

Mating type	*AA*×*AA*	*AA*×*Aa*	*AA*×*aa*	*Aa*×*Aa*	*Aa*×*aa*	*aa*×*aa*
Fertility	7.5	8	5	6	4	3.9

c. Family selection:

	Offspring Fitnesses		
Parental Mating Type	*AA*	*Aa*	*aa*
AA × *AA*	1.05		
AA × *Aa*	0.8	1.04	
AA × *aa*	—	1.9	
Aa × *Aa*	1.5	2.0	1.6
Aa × *aa*	—	1.5	1.4
aa × *aa*	—	—	1.6

4. Suppose a human population in a malarial region is at equilibrium for the sickle cell polymorphism with $p = 0.89$, $q = 0.11$ (the frequencies of the A and S alleles, respectively), the population is randomly mating, and the fitnesses are: *AA*, 0.9; *AS*, 1; and *SS*, 0.2. Now suppose the system of mating changes such that females choose males as mates that maximize the fitness of their offspring. For each of the three female genotypes, calculate the average fitness of their progeny when they mate with each of the three types of males. What male maximizes the average fitness of the female's progeny in each case? Should females ever choose as mates the males with the highest fitness?

5. An allele A shows biased gene conversion over allele a such that in heterozygotes there is a 2 : 1 segregation in favor of A in heterozygotes. Assume for now that this is the only phenotype associated with A.

a. Starting with an allele frequency of 0.1 for A in a random-mating population, what is the frequency of A in the next generation?

b. Same as part a, but now assume nonrandom mating with $f = 0.5$.

c. Assume random mating again but also assume that *AA* homozygotes are lethal but *Aa* and *aa* genotypes are viable and equally fit. Starting with the frequency of *Aa*'s at 0.18 in the adult population, what is the frequency of A in the next generation of adults?

6. Consider the following frequency-dependent fitness model:

Genotype	AA	Aa	aa	
Fitness	$1 - tp^2$	$1 - tpq$	$1 - tq^2$	$t > 0$

where q is the frequency of the a allele. Assume random mating.

a. Write down the equation for average fitness as a function of q and find the value of q that maximizes average fitness. (Hint: Remember your calculus and show explicitly that it is a maximum.)

b. Write down the equation for the average excesses of the A and a alleles and find the polymorphic equilibrium q value. (Hint: The average excesses for both alleles are the same, namely 0, only at equilibrium. Also, remember about factoring polynomials and solving for roots.)

c. Does the equilibrium q value maximize average fitness and is it stable (look at Δq at ± 0.1 from q_{eq})?

d. Calculate the fitnesses of all genotypes at equilibrium. What would you infer about the presence of natural selection at this locus given only the genotypic fitnesses at equilibrium?

7. Given the following fitnesses, show whether or not the polymorphism is protected. If the polymorphism is not protected, identify the alleles that are not protected. In all cases, assume random mating.

a. Competitive interactions:

Genotype	AA	Aa	aa
AA	0.9	0.9	1.2
Aa	0.95	1	1.7
aa	1.2	1.7	1.6

b. Fertility:

Mating type	$AA \times AA$	$AA \times Aa$	$AA \times aa$	$Aa \times Aa$	$Aa \times aa$	$aa \times aa$
Fertility	7	6	5	6	4	5

c. Family selection:

	Offspring Fitnesses		
Parental Mating Type	AA	Aa	aa
$AA \times AA$	1		
$AA \times Aa$	1	1.05	
$AA \times aa$	—	1.9	
$Aa \times Aa$	1.5	2.0	1.6
$Aa \times aa$	—	1.6	1.6
$aa \times aa$	—	—	1.5

8. Consider the following two-locus fitness model:

Genotype at locus *1**2*	*AA*	*Aa*	*aa*
BB	1.2	0.7	1.9
Bb	0.3	1	0.8
bb	0.9	0.4	1.1

Given that the initial gene pool consists only of the *AB* and *ab* gametes each at equal frequency, what is the change under random mating in an infinitely large population of the *AB* gamete for a recombination frequency of 0, for a recombination frequency of 0.025, and for a recombination frequency of 0.1? Does the genetic architecture (in this case as measured by r) affect the initial evolutionary direction in this case?

9. For some systems, there are interactions between diploid and haploid phases (as for some self-sterility alleles in plants). As a result, frequency-dependent selection can be induced that depends upon gamete rather than genotype frequencies. Consider the following gamete frequency-dependent fitness model:

Genotype	*AA*	*Aa*	*aa*	
Fitness	$1-s+tq$	1	$1+s-tq$	$s, t > 0$

where q is the frequency of the a allele. Assuming random mating, write down the equation for average fitness and find the q value that maximizes average fitness. (Hint: Remember your calculus.) Write down the equation for the average excesses of the A and a alleles and find the polymorphic equilibrium q value. (Hint: The average excesses for both alleles are the same, namely 0, only at equilibrium.) Is the equilibrium q value stable and does it maximize average fitness? Calculate the fitnesses of all genotypes at equilibrium. What would you infer about the presence of natural selection at this locus given only the genotypic fitnesses at equilibrium? Suppose that $s = 0.25$, $t = 1$, and the initial q starts at a frequency of 0.3. Calculate the change in q and average fitness from this point. Does this example obey Fisher's fundamental theorem of natural selection?

Answers

1. a. $k = \frac{4}{5} = 0.8$; $\Delta p = G_{Aa}(k - \frac{1}{2}) = 2(0.05)(0.95)(0.8 - 0.5) = 0.029$, $p' = p + \Delta p = 0.079$

b. $k = \frac{4}{5} = 0.8$; $\Delta p = G_{Aa}(k - \frac{1}{2}) = 2(0.05)(0.95)(0.5)(0.8 - 0.5) = 0.014$, $p' = p + \Delta p = 0.064$

c. $p' = p + G_{Aa}(k - \frac{1}{2}) = 0.05 + 0.029 = 0.079$, $\bar{W} = 0.9938$, $a_A = -0.010$, so $\Delta p' = -0.006$. From part a, $\Delta p = G_{Aa}(k - \frac{1}{2}) = 0.029$, so Δp(total) $= -0.006 + 0.029 = 0.023$, $p'' = p + \Delta p = 0.073$.

2. a. From equation 13.12, it is 10^{-5}.

b. Using equation 13.13, first note that $\alpha = 10^{-4} > 1/(2N) = 5 \times 10^{-5}$, so the time to fixation is $4N = 40,000$ generations.

c. Orthologous fixation (whose coalescence time is also $4N$).

3. a. When *A* is rare, the population consists mostly of *aa* homozygotes interacting with one another, so $\bar{W} \approx 1.6$. The *A* allele is found almost exclusively in *Aa* individuals that interact primarily with *aa*'s. So, when *A* is rare, the average excess of *A* is approximately $1.7 - 1.6 = 0.1 > 0 \Rightarrow A$ will increase in frequency. Therefore, *A* is protected.

When *a* is rare, the population consists mostly of *AA* homozygotes interacting with one another, so $\bar{W} \approx 0.95$. The *a* allele is found almost exclusively in *Aa* individuals that interact primarily with *AA*'s. So, when *a* is rare, the average excess of *a* is approximately $0.9 - 0.95 = -0.05 < 0 \Rightarrow a$ will decrease in frequency. Therefore, *a* is not protected. *The polymorphism is not protected.*

(Interesting aside, note that fixation of *A* is stable, resulting in a population with average fitness of 0.95, whereas fixation of *a* is not stable, preventing the evolution of a population with average fitness of 1.6.)

b. When *A* is rare, the population consists mostly of *aa* × *aa* mating pairs with a fertility of 3.9, so $\bar{W} \approx 3.9$. The *A* allele is found almost exclusively in *Aa* × *aa* mating pairs with fertility 4. So, when *A* is rare, the average excess of *A* is approximately $4 - 3.9 = 0.1 > 0 \Rightarrow A$ will increase in frequency. Therefore, *A* is protected.

When *a* is rare, the population consists mostly of *AA* ×*AA* mating pairs, so $\bar{W} \approx 7.5$. The *a* allele is found almost exclusively in *AA* × *Aa* mating pairs with fertility 8. So, when *a* is rare, the average excess of *a* is approximatley $8 - 7.5 = 0.5 > 0 \Rightarrow a$ will increase in frequency. Therefore, *a* is protected. *The polymorphism is protected.*

c. When *A* is rare, the population consists mostly of *aa* × *aa* mating pairs with *aa* offspring of fitness 1.6, so $\bar{W} \approx 1.6$. The *A* allele is found almost exclusively in *Aa* offspring of *Aa* × *aa* mating pairs with fitness of 1.5. So, when *A* is rare, the average excess of *A* is approximately $1.5 - 1.6 = -0.1 < 0 \Rightarrow A$ will decrease in frequency. Therefore, *A* is not protected.

When *a* is rare, the population consists mostly of *AA* × *AA* mating pairs with *AA* offspring of fitness 1.05, so $\bar{W} \approx 1.05$. The *a* allele is found almost exclusively in *Aa* offspring of *AA* × *Aa* mating pairs with fitness of 1.04. So, when *a* is rare, the average excess of *a* is approximately $1.04 - 1.05 = -0.01 < 0 \Rightarrow a$ will decrease in frequency. Therefore, *a* is not protected. *The polymorphism is not protected.*

4.

PROBABILITY OF OFFSPRING GENOTYPES

	AA	*AS*	*SS*		
Fitness	0.9	1	0.2		
Mating				Average Fitness	
AA× *AA*	1	0	0	0.9	
AA× *AS*	0.5	0.5	0	0.95	
AA× *SS*	0	1	0	1	maximum
AS× *AA*	0.5	0.5	0	0.95	maximum
AS× *AS*	0.25	0.5	0.25	0.775	
AS× *SS*	0	0.5	0.5	0.6	
SS× *AA*	0	1	0	1	maximum
SS× *AS*	0	0.5	0.5	0.6	
SS× *SS*	0	0	1	0.2	

No female should ever choose the *AS* males, who have the highest fitness.

5. a.

Genotype	*AA*	*Aa*	*aa*
Starting population	0.01	0.18	0.81

$$p' = p + (\text{frequency of } Aa)(k - \tfrac{1}{2}) = 0.13$$

b.

Genotype	*AA*	*Aa*	*aa*
Starting population	0.055	0.09	0.855

$$p' = p + (\text{frequency of } Aa)(k - \tfrac{1}{2}) = 0.115$$

c.

Genotype	*AA*	*Aa*	*aa*	
Starting population	0.0	0.18	0.82	$p = 0.09$

After conversion: $p' = p + (\text{frequency of } Aa)(k - \frac{1}{2}) = 0.09 + (0.18)(0.1667) = 0.12$

Zygotic frequency	0.0144	0.2122	0.7744	
Fitness	0	1	1	$\bar{w} = 1 - (0.12)^2 = 0.9856$
After lethal selection	0.0	0.2143	0.7857	

$$p'(\text{in adults}) = (0.5)(\text{adult frequency of } Aa) = 0.1071$$

6. a.

$$\bar{w} = p^2(1 - tp^2) + 2pq(1 - tpq) + q^2(1 - tq^2) = 1 - t(p^4 + 2p^2q^2 + q^4)$$
$$= 1 - t(p^2 + q^2)^2 = 1 - t(1 - 2pq)^2 = 1 - t(1 - 2q + 2q^2)^2$$

To find the maximum, take the first derivative as $-2(-2 + 4q)(1 - 2q + 2q^2)t$ using the chain rule from introductory calculus (this can also be found by taking the derivative of the first form of $\bar{w}$ given above and using the chain rule to find the derivative; there is no need to expand the function out entirely). Next, set the first derivative to zero and solve. The only real solution is $\frac{1}{2}$.

To confirm that this is a maximum and not a minimum, take the second derivative as $-2(-2 + 4q)^2t - 8(1 - 2q + 2q^2)t$ and evaluate at $q = \frac{1}{2}$ to be $-4t < 0$, which implies it is a maximum.

b. Average excesses of the A and a alleles:

$$a_A = p(1 - tp^2) + q(1 - tpq) - \bar{w} \qquad a_a = p(1 - tpq) + q(1 - tq^2) - \bar{w}$$

At equilibrium, $a_A = a_a$, which implies $p(1 - tp^2) + q(1 - tpq) = p(1 - tpq) + q(1 - tq^2)$, or $(-1 + 4q - 6q^2 + 4q^3) = 0$.

$(-1 + 4q - 6q^2 + 4q^3) = (-1 + 2q)(1 - 2q + 2q^2)$, so the roots are $\frac{1}{2}$ and $\frac{1}{2} \pm \frac{i}{2}$. The only real root is $q_{\text{eq}} = \frac{1}{2}$.

c. $q_{\text{eq}} = \frac{1}{2}$ does maximize $\bar{w}$ in this case. For stability, look at Δq at ± 0.1 from q_{eq}, that is, $q = 0.4$ and $q = 0.6$.

At $q = 0.4$, $\bar{w} = 1 - (0.2704)t$ and $a_a = (0.0624)t > 0$, which implies q is increasing under selection.

At $q = 0.6$, $\bar{w} = 1 - (0.2704)t$ and $a_a = -(0.0416)t > 0$, which implies q is decreasing. Therefore, q_{eq} is *stable*.

d. Fitnesses of all genotypes at equilibrium:

Genotype	AA	Aa	aa
Fitness	$1 - tp^2 = 1 - \frac{1}{4}t$	$1 - tpq = 1 - \frac{1}{4}t$	$1 - tq^2 = 1 - \frac{1}{4}t$

All genotypes have the same fitness at equilibrium, which creates the false impression that natural selection is not operating at this locus.

7. a. When a is rare, $w_2 \approx 0.9 + 2q(0.95) \approx \bar{w}$ and $w_1 \approx w_{12} + 2qw_{11} = 0.9 + 2q(1)$, so $a_a \approx 0.9 + 2q - 0.9 - 1.9q = 0.1q > 0$, which implies that a is protected.

When A is rare, $w_1 \approx w_{10} = 1.7$ and $w_0 \approx w_{00} = 1.6 \approx \bar{w}$, so $a_A \approx 0.1 > 0$, which implies A is protected.

Both alleles are protected; therefore this is a protected polymorphism.

b. When a is rare, $W_{AA} \approx w_1 = 7 \approx \bar{w}$ and $W_{Aa} \approx w_2 = 6$; so $a_a \approx -1 < 0$ and a is not protected.

When A is rare, $W_{aa} \approx w_6 = 5 \approx \bar{w}$ and $W_{Aa} \approx w_5 = 4$, so $a_A \approx -1 < 0$ and A is not protected.

Therefore, there is no protected polymorphism.

c. When a is rare, $W_{AA} \approx w_{21} = 1 \approx \bar{w}$ and $W_{Aa} \approx w_{12} = 1.05$, so $a_a \approx 0.05 > 0$ and a is protected.

When A is rare, $W_{aa} \approx w_{06} = 1.5 \approx \bar{w}$ and $W_{Aa} \approx w_{15} = 1.6$, so $a_A \approx 0.1 > 0$ and A is protected. Both alleles are protected; therefore this is a protected polymorphism.

8. The initial population is always as follows:

Genotype	$AABB$	$AaBb$	$aabb$	
Fitness	1.2	1	1.1	$\bar{W} = 1.075$
Genotype frequency (Hardy–Weinberg)	0.25	0.5	0.25	

$D = 0.25$

Average excess	AB	ab
	0.025	−0.025

The equation for change in the frequency of AB is always (from equation 13.1)

$$\Delta g_{AB} = \frac{g_{AB}}{\bar{w}} a_{AB} - rD\frac{w_{AB\,/\,ab}}{\bar{w}}$$

$$\Delta g_{AB} = \frac{0.5}{1.075} - r(0.25)\frac{1}{1.075}$$

so for $r =$	0	0.025	0.1
$\Delta g_{AB} =$	0.0116	0.0058	−0.012

Note that r does affect the initial evolutionary direction (+ for $r = 0$, 0.025; − for $r = 0.1$).

9. $\bar{W} = (1-q)^2(1-s+tq) + 2(1-q)q + q^2(1+s-tq) = 1 - s + q(t+2s) - 2tq^2$. To find the maximum (or minimum), $d\bar{W}/dq = t + 2s - 4tq = 0 \Rightarrow q = (t+2s)/(4t)$. To show it is the maximum, $d^2\bar{W}/dq^2 = -4t < 0 \Rightarrow$ a maximum.

$a_A = p(1 - s + tq) + q - \bar{W} = -qs + tq^2; a_a = p + q(1 + s - tq) - \bar{W} = -q(s + t) + tq^2 + s$

At equilibrium, $a_A = a_a = 0 \Rightarrow p(1 - s + tq) + q = p + q(1 + s - tq) \Rightarrow q_{\text{eq}} = s/t$.

To check for stability, note that $a_A = -qs + tq^2 = q(-s + tq)$. Hence, the sign of a_A (and thus automatically a_a) is determined solely by $-s + tq$. If $q > s/t$, then $a_A > 0 \Rightarrow q$ will decrease. If $q < s/t$, $a_A < 0 \Rightarrow q$ will increase. Hence, the equilibrium is stable.

At equilibrium, $w_{AA} = 1 - s + t(s/t) = 1; w_{Aa} = 1; w_{aa} = 1 + s - t(s/t) = 1 \Rightarrow$ *appears* neutral at equilibrium.

At $q = 0.3$, $s = 0.25$ and $t = 1$; $\bar{W} = 1.02$; $a_a = (0.3)(1 + 0.25 - 0.3) + 0.7(1) - 1.02 = -0.0350$.

Hence $\Delta q = qa_a/\bar{W} = (0.3)(-0.035)/1.02 = -0.0103 \Rightarrow q' = q + \Delta q = 0.2897 \Rightarrow \bar{W}' = 1.0167$.

Hence, $\Delta\bar{W} = -0.0033 < 0 \Rightarrow$ Fisher's Fundamental Theorem of Natural Selection (FFTNS) is violated.

CHAPTER 14

Problems

1. Consider a two-niche Levene model with a single locus and two alleles, with random mating at the total population level. Let $w_1 = v_2 = 1.05$ and $w_2 = v_1 = 0.95$. Assume $c_1 = c_2 = 0.5$.

a. Will the polymorphism be protected?

b. Let w_1 =1.06, with all other fitnesses unchanged. Will the polymorphism be protected?

c. Assume the original fitnesses, but let $c_1 = 0.53$ and $c_2 = 0.47$. Will the polymorphism be protected?

d. Assume a hard-selection model by regarding all the c's as z's in the sense of equation 14.6. Redo parts a through c under this assumption.

e. Assume that the population is no longer randomly mating at the total population level but rather that the niches define a subdivided population with $m = 0.1$. Redo parts a through d under this assumption.

2. A population fits the assumptions of a two-niche, soft-selection Levene model with a single locus and two alleles, with random mating at the total population level. Let the fitnesses be:

	Fitness			
Niche	*AA*	*Aa*	*aa*	c_i
1	1.5	1	0.5	0.25
2	0.85	1	1	0.75

- **a.** Will the polymorphism be protected?
- **b.** Assume that selection is hard and that the c's in the above table now correspond to z's (proportion of zygotes entering the niche). Will the polymorphism be protected?
- **c.** Assume that $c_1 = c_2 = 0.5$. Redo part a with this new assumption.
- **d.** Assume that $z_1 = z_2 = 0.5$. Redo part b with this new assumption.

3. Assume an organism is continuously distributed across a space in which the environment changes over a distance of Δ. Suppose the population is polymorphic for a single locus with two alleles, A and a, with a fitness model given by equation 14.10 and a gene flow model given by equation 14.11. Calculate the characteristic length of variation in allele frequencies for the following parameters and indicate whether this environmental shift is experienced as an ecotone or gradient with respect to this locus.

d	m	b	Δ
50	0.3	0.0001	500
50	0.3	0.001	500
200	0.8	0.0001	500
200	0.8	0.001	500
500	0.5	0.0001	500
500	0.5	0.001	500

4. Assume an organism lives in a seasonal environment and is polymorphic for an autosomal locus with two alleles, A and a, such that the genotypes have the following fitnesses:

Genotype	AA	Aa	aa
Fitness season 1	0.7	1	1.2
Fitness season 2	1.3	1	0.7

- **a.** Is this polymorphism protected when the organism has two generations per year, with one generation experiencing season 1 and the next generation experiencing season 2? Assume random mating at this locus.
- **b.** Is this polymorphism protected when the organism has one generation per year, with all living individuals experiencing both seasons? Assume random mating at this locus.

5. Suppose a new allele, A, mutates from the ancestral allele a such that the relative fitnesses are 1.02 and 1 for Aa and aa, respectively. Assume an idealized, random-mating population of large size. Further assume that fine-grained environmental variation induces a variance in fitness across Aa individuals, with the variance in the number of offspring of Aa individuals being $1.02 + \sigma_s^2$.

- **a.** What is the probability of survival of the A mutant for the following values of σ_s^2: 0, 0.5, 1, 2?
- **b.** Assume that the fitness of the Aa type is 1.04 with a variance of $1.04 + \sigma_s^2$. Redo part a with these new assumptions.
- **c.** Contrast the results of parts a and b to identify situations in which a new mutant of lower average fitness has a greater probability of survival.
- **d.** Assume that the total population has an ideal size of 30. Redo parts a, b, and c.

Answers

1. Soft selection:

	w Harmonic Mean	v Harmonic Mean	Protected?
a.	0.9975	0.9975	Yes
b.	1.0020	0.9975	No
c.	1.0005	0.9945	No

d. Hard selection:

	w Arithmetic Mean	v Arithmetic Mean	Protected?
a.	1.0000	1.0000	No
b.	1.0050	1.0000	No
c.	1.0030	0.9970	No

e. Soft selection with $m = 0.1$:

	w, m Harmonic Mean	v, m Harmonic Mean	Protected?
a.	0.7500	0.7500	Yes
b.	0.7619	0.7500	Yes
c.	0.7732	0.7282	Yes

Hard selection with $m = 0.1$:

	w Arithmetic Mean	v Arithmetic Mean	ω_1	ω_2	Protected?
a.	0.7975	0.7975	1.045	0.955	Yes
b.	0.8166	0.7975	1.0545	0.955	Yes
c.	0.8233	0.7730	1.0453	0.9547	Yes

2. a.

w Harmonic Mean	v Harmonic Mean	Protected?
0.8000	0.9533	Yes

b.

w Arithmetic Mean	v Arithmetic Mean	Protected?
0.8750	1.0125	No

c.

w Harmonic Mean	v Harmonic Mean	Protected?
0.6667	1.0851	No

d.

w Arithmetic Mean	v Arithmetic Mean	Protected?
0.7500	1.1750	No

3.

d	m	b	Δ	ℓ_c	Ecotone/Gradient
50	0.3	0.0001	500	122.4744871	Gradient
50	0.3	0.001	500	38.72983346	Gradient
200	0.8	0.0001	500	800	Ecotone
200	0.8	0.001	500	252.9822128	Gradient
500	0.5	0.0001	500	1581.13883	Ecotone
500	0.5	0.001	500	500	Ecotone

4. Because there is no dominance or recessiveness, the conditions for both the coarse-grained case (a) and the fine-grained case (b) depend upon the following fitness products (equation 14.19 and Table 14.4): $v_1v_2 = 0.84$ and $w_1w_2 = 0.91$. As both of these products are less than 1, the polymorphism is protected in both cases.

5. a. From equation 14.20:

σ_s^2	Prob(survival)
0	0.039215686
0.5	0.026315789
1	0.01980198
2	0.013245033

b.

σ_s^2	Prob(survival)
0	0.076923077
0.5	0.051948052
1	0.039215686
2	0.026315789

c. The only case in which the mutant with the lower average fitness has a higher survival probability is the case when $\sigma_s^2 = 0$ in part a contrasted to the case of $\sigma_s^2 = 2$ in part b.

d. Use equation 14.22:

Redoing part a:

σ_s^2	Prob(survival)
0	0.04312255
0.5	0.030612056
1	0.020154376
2	0.006836626

Redoing part b:

σ_s^2	Prob(survival)
0	0.077521637
0.5	0.062773753
1	0.048546645
2	0.024052179

Redoing part c: The probability of survival of a mutation with a fitness of 1.04 with $\sigma_s^2 = 2$ is less than the probability of survival of a mutation with a fitness of 1.02 with $\sigma_s^2 = 0$ and $\sigma_s^2 = 0.5$.

CHAPTER 15

Problems

1. The following life history data were gathered by following three genotypic cohorts (no individual lives to age 4):

	COHORT					
	AA		*Aa*		*aa*	
x = age	l_x	$m_x b_x$	l_x	$m_x b_x$	l_x	$m_x b_x$
0	1	0	1	0	1	0
1	0.25	2	0.5	2	0.5	0
2	0.25	2	0.15	2	0.25	4
3	0	0	0	0	0.1	4

Calculate the net reproductive rate and Mathulsian parameter (use the approximation for the Mathulsian parameter rather than solving it implicity) for each genotype.

2. For the situation described in problem 1, assume that the frequency of the A allele is 0.2. Assume that the species is an annual plant with one generation per year, with the ages above corresponding to the ages in months since germination in the spring. All seeds must overwinter before germinating the following spring. Calculate the average excesses of the A and a gametes for the appropriate fitness measures in this situation under random mating and under an inbreeding system with $f = 0.1$. Identify which allele is increasing in frequency due to natural selection or, if the system is at equilibrium, for each system of mating.

3. For the situation described in problem 2, assume that the species is a perennial shrub, with the ages above corresponding to the ages in years since germination. All seeds produced in a given year must overwinter before germinating the following spring. Calculate the average excesses of the A and a gametes for the appropriate fitness measures in this situation under random mating and under an inbreeding system with $f = 0.1$. Identify which allele is increasing in frequency due to natural selection or, if the system is at equilibrium for each system of mating.

4. A population of ageless individuals has a probability of dying of 0.1 in a time interval. The probability of having mated times the expected number of offspring is 0.11 in a time interval.

a. What is the net reproductive rate of these ageless individuals?

b. Suppose a mutation occurs such that the bearers of this mutant allele die at age n but the probability of having mated times the expected number of offspring is 0.12 in a time interval for ages less than n. What value of n results if the mutant and ageless individuals have identical net reproductive rates?

c. When would the mutant described above be favored by natural selection?

Answers

1.

COHORT

	AA	*Aa*	*aa*
R_0	1	1.3	1.4
m	0	0.1875	0.125

2. In this situation of discrete generations of fixed length, R_0 is the appropriate measure of fitness. Hence:

Genotype	*AA*	*Aa*	*aa*	p	f
Frequency	0.04	0.32	0.64	0.2	0
Fitness	1	1.3	1.4	Average $W = 1.352$	
Genotypic deviation	−0.352	−0.052	0.048		

	A	*a*	
Average excess	−0.112	0.028	*a* is increasing in frequency

Genotype	*AA*	*Aa*	*aa*	p	f
Frequency	0.056	0.288	0.656	0.2	0.1
Fitness	1	1.3	1.4	Average $W = 1.3488$	
Genotypic deviation	−0.3488	−0.0488	0.0512		

	A	*a*	
Average excess	−0.1328	0.0332	*a* is increasing in frequency

3. In this situation of overlapping generations and no fixed generation length, the Malthusian parameter is the appropriate measure of fitness. Hence:

Genotype	*AA*	*Aa*	*aa*	p	f
Frequency	0.04	0.32	0.64	0.2	0
Fitness	0	0.1875	0.125	Average $W = 0.14$	
Genotypic deviation	−0.14	0.0475	−0.015		

	A	*a*	
Average excess	0.01	−0.0025	*A* is increasing in frequency

Genotype	*AA*	*Aa*	*aa*	p	f
Frequency	0.056	0.288	0.656	0.2	0.1
Fitness	0	0.1875	0.125	Average $W = 0.136$	
Genotypic Deviation	−0.136	0.0515	−0.011		

	A	*a*	
Average excess	−0.001	0.00025	*a* is increasing in frequency

4. a. From equation 15.13, $R_0 = 1.1$.

b. Use equation 15.19 and set it equal to 1.1:

$$R_0^{'} = mb' \sum_{x=0}^{n-1} (1-d)^x = \frac{mb'}{d}\left[1-(1-d)^n\right] = 1.2[1-0.9^n] = 1.1$$

Solving for n yields $n = 23.585$.

c. It would be favored for any $n > 23.585$.