Solutions to population genetics exercises

There may be some formatting problems (superscripts and subcripts) due to translation to html. A fully proper hard copy will be handed out in class.

1. Rh blood types in humans are determined by a single locus, in which the Rh+ allele is completely dominant. In a study of 400 humans in the Basque region, 230 were found to have Rh+ blood; the remainder were Rh-. What are the frequencies of the + and – alleles in the population? What do you expect the proportions of the 3 genotypes to be under Hardy-Weinberg equilibrium? If the population were genetically tested using restriction fragment length polymorphisms and the frequency of heterozygotes were found to be lower than anticipated, what might be the reason(s)?

Let q be the frequency of the Rh- allele. All Rh- phenotypes are homozygous, so $q^2 = 170/400 = 0.425$. Taking the square-root, q=0.652. The frequency of the Rh+ allele is p = 1-q = .348. Note that the less frequent allele has the more frequent phenotype due to dominance.

At Hardy-Weinberg equilibrium, the genotypes should be present in proportions p^2, 2pq, and q^2. Thus we expect

fr (++) = $(0.348)^2 = 0.121$

fr (+-) = 2(0.348)(0.652) = 0.454, and

fr (--) = 0.425, as already calculated..

If the genotypes differ significantly from these expected frequencies, one or more of the assumptions of the Hardy-Weinberg theorem have been violated, including the following. Selection may be acting to reduce the frequency of one or more genotypes; inbreeding or assortative mating may be reducing the frequency of heterozygotes; drift may be reducing the frequency of heterozygotes and increasing the frequency of one of the homozygotes; or immigration from another population may be contributing genotypes in numbers that skew the expected frequency.

2. In the fly *Drosophila mauritania*, a transposable element called mariner exists at a several loci in the genome. Each copy is deleted from the genome with a frequency of 1 percent per generation. In a population in which mariner is fixed at a specific locus, how many generations will it take until the frequency of individuals that are homozygous for the deletion is 5 percent? Hint: this question involves both Hardy-Weinberg and one-way mutation calculations.

Let p be the frequency of the presence of the intron and q be the frequency of its absence. Let u be the mutation rate, and u=0.01. We know that under one-way mutation,

$p_t = p0e^{-ut}$.

In this case, p0 = 1.00. We want to know t when $p^2 = 0.05$, which occurs when p = 0.223. Plugging into the equation above,

$0.223 = (1.00)\ e^{-(.01)t}$

Taking logarithms of both sizes,

-1.50 = - 0.01t, and

t = 150 generations.

3. Consider a locus with two alleles, A and a. If the mutation rate from A to a is $0.86 * 10^{-3}$ mutations per generation and the mutation rate from a to A is $0.47 * 10^{-3}$, what is the expected equilibrium frequency of a?

Let u = rate of forward mutation = $0.86 * 10^{-3}$, and let v= rate of back-mutation = $0.47 * 10^{-3}$. Let p be the frequency of A and q be the frequency of a. We know that at equilibrium p = v/(u+v), and q = 1-p. So $q = 1- [0.47 * 10^{-3}/(0.86* 10^{-3} + 0.47 * 10^{-3})] = 0.65$. This makes sense, because the rate of mutation making a is faster than the rate of mutation turning it back into A. (Try to always check your answers in this common sense, qualitative kind of way).

4. In a haploid organism, with the same mutation rate, what is the expected equilibrium frequency of a? Suppose there is no back-mutation -- only forward mutation (A to a) at the rate given in #3 -- and A is 10% more fit than a. What is the expected equilibrium frequency of a? Hint: you will have to derive an expression for the equilibrium value, using the same technique we used for the diploid case in class.

The equilibrium expression makes no distinction between haploid and diploid (it doesn't matter how many copies of the gene are in each individual – they all contribute to the population frequency and the rate of mutation affects them all equally). So the expected equilibrium is the same as in the problem above, q= 0.65.

Now we want to know about a selection-mutation equilibrium, so make a fitness table for the haploid case.

A	a
frequency p	q
w 1	1-s
frequency' p/$\bar{w}$	q(1-s)/$\bar{w}$

where frequency' is the frequency in the next generation and $\bar{w}$ is the mean fitness = p + (1-s)q = p+q-sq = 1-sq.

Delta-q (the change in q in each generation due to selection) can be written this way

delta-q = q - (q-sq)/(1-sq)

= $(q - sq^2 - q + sq)/(1-sq)$

= sq (1-q)

= sqp/(1-sq)

Near equilibrium, the mean fitness is very close to 1, so we can drop the denominator and say that

delta-q ~ sqp

We also know that delta-q due to mutation is

delta-q = up

At equilibrium, the rate of change of the allele due to mutation equals the rate of change due to selection, so

sqp ~ up

which simplifies to

q ~ u/s.

Note that this result is the same as the diploid case without the square root, because selection acts on all a's in the haploid, not just on the homozygotes (frequency q^2), as it does in the diploid.

Plugging in the given values, q at equilibrium will equal $(0.86 * 10^{-3})/0.1 = 0.086$, or about 9 percent (a substantial equilibrium frequency for a deleterious allele).

5. Imagine two gene loci, each with two alleles. In both cases, the homozygote AA has 1 percent greater fitness than the homozygote aa, and BB has 1 percent greater fitness than bb. When the experiment begins, both A and B have an allele frequency of 1 percent. If A displays complete dominance over a, and B displays additive codominance with b, which gene will be evolving faster after 100, 500 and 1000 generations? Assume that at these times, the frequencies of A and B respectively are about 0.031 and 0.017; 0.471 and 0.130; and 0.816 and 0.679, respectively.

For a dominant allele,

$delta\text{-}p = spq^2/(1-sq2)$

For a codominant allele,

$delta\text{-}p = sqp/2(1-spq-sq^2)$

Plugging in the respective values s (0.01) and p and q at each time, we get

time	delta-p (A)	delta-p (B)
100	2.94E-04	8.44E-05
500	1.32E-03	5.70E-04
1000	2.76E-04	1.09E-03

This result shows that at low and intermediate frequencies, the dominant is evolving faster because selection on the heterozygote is stronger than in the codominant case. At high freq

6. What is the expression for the expected equilibrium value of an allele frequency (p) under the influence of mutation (at rate u) and selection (coefficient of selection = s), if the advantageous allele is codominant? Hint: Adapt what we did in class for a dominant allele for this case.

We know from the lecture that the rate of change of a codominant under selection is $(1/2)sqp/(1-spq-sq^2)$. (If you need to verify this, you can use a fitness table.)

The rate of change of an allele due to mutation is

delta-q = up, and

delta-p = -up.

At equilibrium, the rate due to selection equals the rate due to mutation. We assume that near equilibrium the mean fitness is very close to 1, so

$up = sqp/2$

$u = sq/2$

$q = 2u/s.$

7. Consider a lethal gene for the enzyme phosphodiesterase, with two alleles, PDE and PDE-. PDE is mutated to PDE- at the rate $5 * 10^{-6}$ mutations per generation with no back mutation. If PDE- is completely recessive, what is its equilibrium frequency? What is the equilibrium frequency if the heterozygote is 40 percent less fit than the AA genotype? What assumption did you make to make this calculation? Is it valid? If you had not made this assumption, how would your answer have changed (answer qualitatively -- higher or lower -- don't try to calculate the amount).

Let q be the frequency of PDE-, which in the first case is completely recessive. If the allele is lethal in the homozygote, its fitness is 0, and its selection coefficient, which we will call s = 1. We have already seen that the mutation-selection equilibrium frequency of a recessive is

$q = (u/s)^{1/2}$

Plugging in, $q = (5 * 10^{-6}/1)^{1/2} = .0022$ or 0.2 percent. Notice that this is a small but nontrivial frequency for a LETHAL allele.

For the case of partial dominance, we again want to know the rate of change of p under selection (this is the general case of which partial dominance is a special case). The simplest way to do the calculations is to use a term h that expresses s_2 in terms of s. Thus, a fitness table looks like this:

AA	Aa	aa
frequency p^2	$2pq$	q^2
w 1	$1-hs$	$1-s$
frequency' p^2/w	$2pq(1-hs)/w$	$q^2(1-s)/w$

$w = 1 - 2pqhs - sq^2$

The change of p in one generation due to selection is

$delta\text{-}p = pq/w*[p(hs)+q(1-h)s]$

The change of p due to mutation is

$delta\text{-}p = -up$

At equilibrium, these two equal each other, and the mean fitness is very close to one, so we approximate

$up = pq[phs+sq(1-h)]$

Cancelling p and pulling s out of each term gives

$u = sq[ph+q(1-h)]$

$u = sq[(1-q)h + q(1-h)]$

$u = sq[h-hq+q-hq]$

$u = sq[h+q-2hq]$

$u = sqh[1+q/h-2q]$

$q = u/[sh(1-2q+q/h)]$

The equilibrium value of q will be small if the mutation rate is low compared to the selection coefficient, so the denominator will reduce to approximately

$q = u/sh$

Note that this general case gives the same results that we derived in class for additive codominance (h = 0.5). You can now use it for partial dominance. It does not work as h gets closer to zero (total dominance), because then the denominator above cannot be reduced to sh. In this case,

$q = (5*10^{-6})/(1)(0.4) = 1.25*10^{-5}$

Note that the equilibrium frequency of the deleterious allele is much lower than for complete dominance, because partial dominance allows the allele to be selected against even in the heterozygous state.

[This part of the problem was probably too hard for me to include here. Don't worry if it seems difficult to do.]

8. Imagine a small diploid population of 10 individuals. Consider a gene for an enzyme that has two alleles (A and a), which begin at equal frequency and are selectively equivalent. What is the probability that A will be completely lost from the population in the next generation? About how many generations will it take before A has frequency 0.1? Hint: the expected value of p at time t is given as $ptqt/p0q0 = (1-1/2N)^t$.

At time t=0, p=q=0.5. The probability that the next generation will contain any specific number of alleles *i* is given as

$Pr\ (i) = 2N!/[2N-i)!(i!)]p^iq^{2N-1}$

Here, we want to know the chance that the number of alleles i will = 0, and N=20. The probability of loss therefore equals

$Pr\ (O) = 2N!/[(2N!)(1)](0.5)^0*(0.5)^{20}$

All terms here cancel except the last, leaving

$Pr\ (0) = (0.5)^{20} = 9.5 * 10^{-7}$.

So the chance of loss in the first generation is very small.

Next we want to know how long it will take for p = 0.1. The hint tells us that

$ptqt/p0q0 = (1-1/2N)^t$

We know that when x is small, $(1-x)^t = e^{-xt}$, so

$ptqt/p0q0 = e^{-t/2N}$

Plugging in,

$(0.1)(0.9)/(0.5)(0.5) = e^{-t/2N}$

$= 0.09/0.25 = 0.36 = e^{-t/2N}$

Taking the logarithm of both sides,

$-1.021 = -2Nt = -t/20$

$t = 20.4$

So it will take 21 generations before p reaches 0.1.

9. Imagine a large population split into 50 subpopulations (demes), each with 20 individuals in it. All populations begin with the frequency of A at 0.3. After 1000 generations, how many demes do you expected to be fixed for A? In how many will A have been lost completely? After 30 generations, what do you expect the proportion of heterozygotes in the total population to be? (See hint for #8.)

1000 generations is long enough for virtually all populations to reach fixation for one allele or the other. The probability of fixation of any allele equals its initial frequency in the population, so the probability of any population being fixed for A is 0.3, and the total number of populations expected to be fixed for A is 50 * 0.3 = 15. We expect A to have been lost in the other 35 populations.

After 30 generations, we know that

$ptqt/p0q0 = (1-1/2N)^t$

Plugging in,

$ptqt/(0.3((0.7) = (1-1/40)^{30}$

$ptqt/0.21 = .468$

$ptqt = .098$

We expect the proportion of heterozygotes to = 2pq, so we just multiply the expression above by two, and get the answer that the expected proportion of heterozygotes in any of these populations after 30 generations is 19.6 percent.

10. Consider a diploid population of a plant species that can either self or mate with other individuals with equal probabilities. The population begins with 30 unrelated individuals. What is the probability that any one individual will be autozygous (inbred, with gametes derived from the same parent) in the first generation of offspring) in the first generation? On average, how many individuals in the population will be autozygous in the first generation? What is the probability that any one individual will be autozygous after 100 generations?

The probability of inbreeding in one generation is

$F = 1/2N = 1/60$, or about 0.017.

We expect then that the number of inbred individuals in the first generation will be 1/60 * 30, or 0.5, which means that on average one individual will be inbred in half the populations in which this "experiment" occurs.

The probability of inbreeding after t generations is

$F = 1 - (1-1/2N)^t$

In this case,

$F = 1 - (59/60)^{100} = 1 - 0.186 = 0.814$. So after 100 generations, over 80 percent of the individuals are inbred.

11. For the population in #10, consider a two-allele locus with initial frequency p=.3. After 100 generations, what is the change in p due to inbreeding? What is the expected proportion of heterozygous individuals in the population after 100 generations, considering only the effect of inbreeding? What is the expected proportion of heterozygotes at this time due to the effect of drift? What is the expected proportion of heterozygotes at this time when both inbreeding and drift are considered.? (Hint: think about the way that each process affects heterozygosity. Then think how one might compound the other.)

Inbreeding does not change allele frequencies, so the change in p is zero. It still equals 0.3.

F gives the proportional reduction in heterozygotes due to inbreeding as

$$1-F = Ht/H_0.$$

We calculated F for 100 generations and N = 60 above, and we know that the proportion of heterozygotes initially was 2pq = 2(0.3)(0.7)= 0.42. Thus, after 100 generations, the expected proportion of heterozygotes is

$H_t = H_0(1-F) = 0.42*(1-0.814) = 0.078.$

So only about 8 percent of the population will be heterozygous, due solely to inbreeding.

We can calculate the proportion of heterozygotes due to drift using the expression from the hint above:

$$ptqt/p0q0 = (1-1/2N)^t$$

Filling in,

$ptqt = (0.21)(59/60)^{100} = .039$

The proportion of heterozygotes is 2pq, so H = 2(0.039) = 0.078. Note that the reduction in heterozygotes due to drift is equal to that due to inbreeding, but the causes (change in allele frequency in one case, chance union of IBD gametes in the other) are different.

Drift and inbreeding will occur in small populations simultaneously. Since inbreeding is a reduction in heterozygotes above and beyond what is expected based on the allele frequencies (2pq), we can calculate the total loss of heterozygosity by first examining the reduction due to changing allele frequencies (drift) and then calculate inbreeding as the additional loss of heterozygotes on top of drift.

We have already calculated the expected proportion of heterozygotes due to changing allele frequencies as

$2ptqt = 0.078$

Inbreeding will reduce heterozygosity below this expectation by the inbreeding coefficient F, which we already calculated. Thus H_t due to both drift and inbreeding will be given as

$H_t/H_0 = (1-F)(2p_tq_t) = (1 - 0.814)(0.078) = 0.015.$

Since $H_0 = 0.42$, we can say that $H_t = 0.42*0.015 = 0.006$.

We see that although inbreeding or drift alone reduced heterozygosity by about 80 percent, leaving about 8 percent of the population as heterozygotes, the action of the two together leaves less than 1 percent of the population as heterozygotes.

12. Consider two populations of mountain goats separated by a high pass that is snowed over most of the year and difficult to get over. In these goats there is a gene (B, b) for which the frequencies are 0.3 and 0.7. If the populations have 100 goats each — of which 2 are migrants at each generation — what proportion of the population is expected to be heterozygous? What is the equilibrium heterozygosity if the populations have 1000 goats each, of which 2 are migrants in each generation?

We know that F at equilibrium = 1/(1+4Nm). There are two migrants at each generation, so F = 0.11. Note that the population size does no matter.

F is the proportional reduction in heterozygosity, and the original heterozygosity in the population was 2pq = 0.42, so

$H_t = (1-F)H_0 = 0.89 * 0.42 = 0.373$

About 37 percent will be heterozygous at equilibrium.