Name: **KEY**

Student number:

**MID-TERM BIO 434: October 25, 2001**

<table>
<thead>
<tr>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
</tr>
<tr>
<td>Q2</td>
</tr>
<tr>
<td>Q3</td>
</tr>
<tr>
<td>Q4</td>
</tr>
<tr>
<td>Q5</td>
</tr>
<tr>
<td>Q6</td>
</tr>
<tr>
<td>Q7</td>
</tr>
<tr>
<td>Q8</td>
</tr>
</tbody>
</table>
1. (12 points) What evolutionary processes can increase or maintain genetic variation in populations? What decreases genetic variance? Why?

Genetic variance is increased or maintained by processes which move allele frequencies away from zero or one. These include mutation, overdominance, balancing selection, and negative frequency dependent selection.

Genetic variance is reduced by processes which act to remove alleles from populations. This includes directional selection, genetic drift, underdominance, etc.
2. A population of fritillary butterflies has 20 males and 30 females. Within a sex, each individual is equally likely to be the parent of any offspring. By gel-electrophoresis, the allele frequency for the faster moving allozyme at an isocitrate dehydrogenase (IDH) locus was 30% in males and 50% in females. Assuming that there are no selective differences between the two alleles at this IDH locus and that the locus is autosomal:

a. (4 points) What is the allele frequency for the fast allele in this population?

The frequency is \( \frac{1}{2} \times 0.3 + \frac{1}{2} \times 0.5 = 0.4 \)

b. (4 points) On average, what should we expect the heterozygosity to be after 10 generations at this locus?

\[
N_e = \frac{4 \times 20 \times 30}{20 + 30} = 48
\]

\[
H_0 = 2 \times 0.4 \times 0.6 = 0.48
\]

\[
H_{10} = \left(1 - \frac{1}{2 \times 48}\right)^{10} \times 0.48 = 0.432
\]

c. (4 points) On average, what should we expect the allele frequency to be at this locus after 10 generations?

These alleles are neutral, so on average the allele frequency isn't changed from one generation to the next. Therefore the answer is 0.4.

d. (4 points) After an infinite number of generations, what is the probability that this fast allele has been lost from the population?

The probability of loss is 1 - the probability of fixation

The probability of fixation is equal to the initial allele frequency for neutral alleles.

\[ 1 - 0.4 = 0.6 \]
3. (10 points) How could we estimate the effective population size from a natural population?

There at least three possibilities:

1. You can estimate demographic quantities of the populations (such as the census size, the variance in reproductive success, the change in populations size over generations) and use the $N_e$ equations to predict $N_e$.

2. You can estimate heterozygosity of neutral markers, then with an estimate of the mutation rate use the heterozygosity to solve for $N_e$ using $H = \frac{4N_e\mu}{1 + 4N_e\mu}$.

3. You could follow changes in heterozygosity of neutral markers over time, and estimate the $N_e$ from the rate of change of heterozygosity.
4. Two sets of bacterial populations were created in the lab: one set used a strain of bacteria with a high mutation rate while the other set used a low-mutation strain. Each population was founded by a single bacterium and allowed to grow to carrying capacity.

a. (5 points) Both sets were introduced to a new environment. Will the change in mean fitness of these populations be higher or lower in the populations founded by bacteria with high mutation rates? Why?

The high mutation population will have a higher genetic variance, and it could therefore adapt to the new environment more quickly. Therefore it is expected to have a higher mean fitness than the low mutation population.

b. (5 points) If these bacteria were introduced into an environment to which it was well adapted and kept at a very small population size, would the change in mean fitness be higher or lower with an increased mutation rate? Why?

In an environment that the bacteria are well adapted to, a very small population will be likely to fix deleterious alleles. In a population with a higher mutation rate, the number of fixed deleterious alleles will be higher, and the high mutation population will have a lower mean fitness.
5. (2 points each: 20 points) Compare a population with a high effective population size to one with a smaller effective size, for the expectations of the following properties, *all else being equal*. (Insert a greater than $>$, less than $<$, or equal sign between the two columns to indicate which is expected to be highest for the given property.)

<table>
<thead>
<tr>
<th>Smaller $N_e$</th>
<th>$&gt;$, $&lt;$, or $=$?</th>
<th>Larger $N_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic variance within populations</td>
<td>$&lt;$</td>
<td>Genetic variance within populations</td>
</tr>
<tr>
<td>Heterozygosity</td>
<td>$&lt;$</td>
<td>Heterozygosity</td>
</tr>
<tr>
<td>Mean fitness</td>
<td>$&lt;$</td>
<td>Mean fitness</td>
</tr>
<tr>
<td>Variance in reproductive success</td>
<td>$&gt;$</td>
<td>Variance in reproductive success</td>
</tr>
<tr>
<td>Bias in sex ratio</td>
<td>$&gt;$</td>
<td>Bias in sex ratio</td>
</tr>
<tr>
<td>Probability of fixing a beneficial allele</td>
<td>$&lt;$</td>
<td>Probability of fixing a beneficial allele</td>
</tr>
<tr>
<td>Probability of fixing a deleterious allele</td>
<td>$&gt;$</td>
<td>Probability of fixing a deleterious allele</td>
</tr>
<tr>
<td>Genetic variance among populations</td>
<td>$&gt;$</td>
<td>Genetic variance among populations</td>
</tr>
<tr>
<td>Substitution rate</td>
<td>$=$</td>
<td>Substitution rate</td>
</tr>
<tr>
<td>Number of alleles</td>
<td>$&lt;$</td>
<td>Number of alleles</td>
</tr>
</tbody>
</table>
6. (10 points) Why does overdominance maintain genetic variation? (Please be as mathematical as possible in your answer; if you can't provide a proof, at least describe the reason verbally.)

There are two ways (at least) to go about this. You can prove that the equilibrium allele frequency is \( p = \frac{t}{s+t} \), then show that this value of \( p \) is between 0 and 1, so that genetic variance would be maintained. Alternatively, you could show that the allele frequency would increase when rare and decrease when common, showing that some intermediate allele frequency was stable, therefore maintaining genetic variance.

For the first approach, consult your answers to one of the assignments.

For the second approach, we can find the change in allele frequency from one generation to the next, assuming that the three genotypes have fitnesses 1-s, 1, and 1-t:

\[
p' = \frac{(1-s)p^2 + p(1-p)}{1-p^2s - qt} = p \frac{1-sp}{1-p^2s - qt}.
\]

Therefore this allele frequency will increase if \( \frac{1-sp}{1-p^2s - qt} > 1 \).

When \( p \) is very small, the condition for increase is approximately \( 1/(1-t) > 1 \), which is true when \( t>0 \). So allele frequency will increase when rare.

The change in the frequency of the other allele will be positive if \( \frac{1-tq}{1-p^2s - qt} > 1 \), which is also positive when \( q \) is small.

Therefore either allele will increase when rare with overdominance, and therefore genetic variation is preserved at this locus.
7. If the mutation rate to a deleterious allele from the wild type allele is $10^{-6}$ and the average number of offspring of the three genotypes is 1.2, 0.9, and 0.8, then

a. (6 points) What is the equilibrium frequency of the deleterious allele, assuming no migration in a large population?

<table>
<thead>
<tr>
<th>genotype</th>
<th>fitness</th>
<th>relative fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>1.2</td>
<td>1</td>
</tr>
<tr>
<td>Aa</td>
<td>.9</td>
<td>0.75</td>
</tr>
<tr>
<td>aa</td>
<td>.8</td>
<td>0.666666</td>
</tr>
</tbody>
</table>

\[
\hat{q} = \frac{\mu}{hs} = \frac{10^{-6}}{0.25} = 4 \times 10^{-6}
\]

b. (6 points) If the frequency of the good allele, measured before selection, is 0.9 in one generation, what would we expect the homozygosity to be at this locus before selection in the next generation?

\[
p' = \frac{p^2 + pq(0.75)}{p^2 + 2pq(0.75) + q^2(0.6666)}(1 - \mu)
\]

\[
= 0.922(1 - 10^{-6}) \approx 0.922
\]

Therefore the expected homozygosity is 2 pq = 2 
\[(0.922)(0.078) = 0.144.\]
8. a. (5 points) In studying your favorite bird species, you sequence several copies of a nuclear pseudogene. You determine that the expected heterozygosity of this pseudogene is 0.2%, and you know that the mutation rate per base pair at this gene is $10^{-8}$. Using these data, what is the best estimate of the effective population size for these birds?

\[
H = \frac{4N_e \mu}{1 + 4N_e \mu} = 0.002
\]

\[
N_e = 50100
\]

b. (5 points) If you did a similar analysis using synonymous codons with \textit{E. coli}, would your estimate of $N_e$ be approximately right, too high, or too low? Explain.

In \textit{E. coli}, synonymous codons are not selectively neutral, but there is selection towards a particular codon (this results in codon bias). As a result, the heterozygosity is lower than would be predicted by the neutral model. Therefore using the method above would lead to an underestimate of the effective size of the \textit{E. coli} population.