7.03 Problem Set 6 Answer Key

Due Monday May 4th 2015 by 3pm

1. The Tasmanian devil (*Sarcophilus harrisii*) is afflicted by devil facial tumor disease (DFTD), a parasitic cancer that is highly contagious and has decimated the population. Researchers working for the Australian government have identified a gene (TAZ1) that seems to control susceptibility to DFTD. The gene is X-linked, and the susceptibility mutation is recessive. Out of 2,000 male devils, 10 have been found to have the susceptibility mutation.
(note: assume sex determination in Tasmanian devils works the same as in humans)

a) What are the frequencies (*p* and *q*) of the major and minor alleles?

b) You approximate that this mutation arises at a rate of $4x10^{-4}$/generation. What is the selective disadvantage of this mutation, assuming we are at steady state?

A new strain of DFTD cells spreads over the Tasmanian devil population. Detailed analysis shows that the above mutants are yet more susceptible to this strain; the mutants' **fitness** is now 0.15.

c) Assuming that this strain is now infecting the same Tasmanian devil population, what are the new allele frequencies (*p*' and *q*') after one generation under the new strain, assuming we are no longer at steady state?

d) After many generations with the new strain, the population equilibrates to 1,500 male devils, 30 of which are highly susceptible to DFTD. Given the same fitness as in part (d), what is the actual mutation rate to this allele?

Further study of TAZ1 finds that its product forms a complex with the product of another gene, which for simplicity's sake we will call TAZ2. A particular mutation in TAZ2 initially seems to confer a completely identical phenotype with respect to DFTD susceptibility when two mutant copies of the gene are present. However, it affects males and females equally.

f) Analysis of the total Tasmanian devil population of both sexes finds that of 3,000 devils, 75 of them are highly susceptible to DFTD due to the TAZ2 mutation. What is the rate of mutation to this allele?

g) As it turns out, devils with only one copy of this particular TAZ2 mutation are actually somewhat resistant to DFTD. Under a model in which the heterozygote advantage perfectly balances the deleterious effect of TAZ2 homozygotes, what are the selective advantage and fitness of these heterozygotes?

h) What does the presence of a heterozygote advantage suggest about your estimate of the mutation rate in part (f)? Explain in two sentences or less.

2. In an effort to understand animal behavior, a colleague of yours has been studying a population of meerkats on an island off the coast of Namibia. This population is now completely isolated, given that meerkats cannot swim the distance from the mainland that would be required to reach the island. She notices that some of the meerkats die in infancy, and since there do not appear to be any major predators that selectively target young meerkats on the island, she suspects a genetic cause. In general, meerkats in the wild avoid mating with kin, but she suggests that due to the isolation of this population, such a taboo may have been abandoned out of necessity.

She observes that in captivity, where breeders are very careful to avoid mating related meerkats, the frequency of infant death is 1 in 100. On this island, the frequency of infant death is 8 in 100.

a) Your colleague is able to collect DNA samples from many meerkats **from the same family** who have died in infancy. Your initial sequencing analysis suggests that there are a number of recessive mutations that seem to lead to this infant death. If these meerkats will only mate with second cousins (sharing the same great-grandparents), how many separate mutations would you expect to find **in this family**, given you have enough sequence data from many offspring of second cousins?

b) If you were able to sequence the great-grandparents of this family of meerkats on this island, how many of these recessive mutations would you expect each one to be a carrier for?

c) You actually find that there are 10 recessive mutations that are responsible for all of the (genetically linked) infant meerkat death in this family. What does this mean about your hypothesis about inbreeding in these meerkats? Explain briefly, **no calculation required.**

3) *Papilio memnon* is a species of swallowtail butterfly found commonly in Southeast Asia. Butterflies of this species have evolved to closely mimic the appearance of other species that are unpalatable to birds of prey. This complex mimicry is controlled by a number of loci that display linkage disequilibrium and have been studied for many years. You have identified a new bi-allelic locus that you believe is involved in determining the forewing color of female *Papilio memnon* and wish to study if this locus is at linkage disequilibrium with a known mimicry gene. [8 PTS]

Your new forewing color gene has two alleles: F and f
The known hindwing pattern gene has two alleles: H and h

You collect 520 butterfly gametes and observe the following haplotype frequencies:

Gamete	Gamete #	Gamete Frequency
FH	312	
Fh	52	
fH	104	
fh	52	

a) Using a statistical test, determine if these genes are in linkage equilibrium. Please state your null and alternative hypotheses and show all your work. Fill in the table. [2 PTS]

Your lab mate informs you that there is another bi-allelic mimicry gene that displays LD with the hindwing pattern gene you used in (a). She suggests that it might be a good control for you to test linkage disequilibrium for this gene with your new forewing color gene. The gene your lab mate suggested has alleles A & a. You find that indeed gene F and gene A are in linkage disequilibrium with statistical significance after observing the following gamete numbers:

Gamete	Gamete #	Gamete Frequency
FA	181	
Fa	55	
fA	273	
fa	37	
Total	546	

b) Fill in the table and calculate the measure of linkage disequilibrium (D) for both sets of data provided. [1 PT]

c) Can you draw any conclusions by comparing your results in part (b)? Why or why not? If not, provide a better method to compare linkage disequilibrium between these two datasets and show any calculations you make. Explain **in one short sentence** why this method is preferred. [2 PTS]

d) Calculate the r^2 measure for both datasets. In **one sentence** explain what r^2 represents. [1 PT]

e) You find that loci H & A (your controls) have an r^2 value of 0.90 with respect to each other however their LD measure is similar to those you have calculated with locus F. In evolutionary terms what could explain this discrepancy?

f) You find that the rate of recombination between the F & H loci is 0.012. Since you are unsure how to determine when linkage disequilibrium first arose in *Papilio memnon* for these loci you decide to use the D calculated from your experiment as a starting point. How many generations would it take for D to decay to a value of 0.01 from its current value? [1 PT]