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?

q = 10/2000 = **0.005**
p = 1-q = **0.995**

[1 pt, 0.5 pt each]

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?

$\Delta q_{sel} = -Sq*(1/3)$
because this allele is X-linked recessive, so 1/3 of the X chromosomes belong to males (assuming equal numbers of males and females)

$\Delta q_{mut} = \mu$
at steady state, $\Delta q_{sel} + \Delta q_{mut} = 0$
$-Sq/3 + \mu = 0$
$S = 3\mu/q$
$S = 3*4x10^{-4}/0.005 =$ **0.24**
[1 pt]

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?

fitness = 1-S -> S = 0.85
$\Delta q_{sel} = -Sq*(1/3) = 0.85*0.005*(1/3) = 0.0014$
$\Delta q_{mut} = 0.0004$

$q' = q - \Delta q_{sel} + \Delta q_{mut} =$ **0.004**
$p' = 1-q' =$ **0.996**

[1 pt]

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?

$q = 30/1500 = 0.02$
$\mu = Sq/3 =$ **5.67×10^{-3}**

[1 pt]

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?

$q = \text{sqrt}(75/3000)$
$\Delta q_{sel} = -Sq^2$

$\Delta q_{mut} = \mu$
at steady state, $\Delta q_{sel} + \Delta q_{mut} = 0$
$-Sq^2 + \mu = 0$
$\mu = Sq^2 = (0.85)(75/3000) =$ **0.02125**

[1.5 pt, 0.5 pt for switch to autosomal recessive]

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?

(note: given that q is so high (by mistake in the problem set), no points were taken off if you did not assume that p is about 1)

$\Delta q_{sel} = \Delta f(a/a) + (\frac{1}{2})f(A/a) = 0$
$-Sq^2 + (1/2)(2hq) = 0$
$hq = Sq^2$

h = Sq = **0.13**
fitness = 1+ h = **1.13**

[1.5 pt]

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.
A heterozygote advantage means that some of the recessive alleles are carried into the next generation by heterozygotes with increased fitness, rather than created by new mutations. Thus, the estimate of mutation rate is too high.

[1 pt]

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?

The probability of two second cousins carrying the same recessive lethal allele from their great-grandparents is 1/64. Then, the probability of their offspring being homozygous for this allele is 1/256 (1/64*1/4). So, each recessive lethal contributes 1/256 to the probability that these offspring will die due to getting two copies of these lethal alleles.

Looking at the statistics on frequency of infant death, it seems that there is an increase of 0.07 (0.08-0.01) of infant death on the island. Assuming that this difference is entirely due to second-cousin inbreeding, to figure out how many recessive lethal alleles are contributing, divide 0.07 by 1/256. $0.07/(1/256) = 17.92 \approx$ **18**.

[2 pt]

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?

Given that 18 alleles appear to be shared by both great-grandparents (on average), each individual one will carry **9** on average.

[1 pt]

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.**

The fact that there are fewer recessive mutations that are linked to infant death suggests that there may be some more closely-related (first cousins or brother-sister) matings that are happening. In this case, some of these mutations are being inherited from grandparents or parents in offspring (so they contribute more strongly to the increased infant death).

[1 pt]

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	0.6
Fh	52	0.1
fH	104	0.2
fh	52	0.1

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]
Using gamete frequencies to calculate allele frequencies,

Alleles	Allele Frequencies
F	0.6 + 0.1 = 0.7
f	0.2 + 0.1 = 0.3
H	0.6 + 0.2 = 0.8
h	0.1 + 0.1 = 0.2

	Observed	Expected	$(O-E)^2$	$(O-E)^2/E$
FH	312	0.7*0.8*520 = 291.2	432.64	1.49
Fh	52	0.7*0.2*520 = 72.8	432.64	5.94
fH	104	0.3*0.8*520 = 124.8	432.64	3.47
fh	52	0.3*0.2*520 = 31.2	432.64	13.9
Total	520	520	χ^2	**24.8**

Df = [(2-1)*(2-1)] = 1

Ho : The H & F loci are in linkage equilibrium
Ha: The H & F loci are in linkage disequilibrium

$p<0.01$, therefore we can reject the null hypothesis, the two genes are in linkage disequilibrium.

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	0.331
Fa	55	0.101
fA	273	0.5
fa	37	0.068
Total	546	

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

For F & H,
$D = pFH_{observed} - pFH_{expected} = 0.6 – (0.7)(0.8) = 0.04$

For F & A,
$D = pFA_{observed} - pFA_{expected} = 0.331 – (0.331+0.101)(0.331+0.5) = -0.0280$

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]

You can not compare the values of D obtained in (b) because the value of D depends on allele frequencies and thus is not comparable across loci. A comparable measure for linkage disequilibrium is D' because it is D normalized to its respective allele frequency.

For F & H,
D>0, find Dmax

$pFh_{expected} = 0.14$, $pfH_{expected} = 0.24$
$pFh_{expected}$ is smaller therefore Dmax = 0.14
D' = D/Dmax = 0.04/0.14 = 0.286

For F & A,
D<0, find Dmin
$-pFA_{expected} = -0.36$
$-pfa_{expected} = -0.096$
$-pfa_{expected}$ is larger therefore Dmin = -0.096
D' = D/Dmin = -0.0280/-0.096 = 0.292

Thus the D' of these two datasets is very close indicating a similar level of linkage disequilibrium.

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

For F & H,
$r^2 = D^2/pF*pf*pH*ph = 0.04^2/(0.7)(0.3)(0.8)(0.2) = 0.048$

For F & A,
$r^2 = D^2/pF*pf*pA*pa = -0.028^2/(0.432)(0.831)(0.568)(0.169) = 0.0227$

r^2 is a measure of allelic correlation and describes how informative the presence of one allele is for the presence of the other in a gamete.

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?
A mutation that arose in the same haplotype with an allele long ago will have a higher r^2 than a mutation that arose in the same haplotype more recently.Here it could be that the A allele arose alongside the H allele a long time ago (or vice versa) and that the F allele arose alongside these much more recently.

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]

$D_n = (1-r)^n * D_0$
$0.01 = (1-0.012)^n * 0.04$
$0.01/0.04 = 0.988^n$
$\ln 0.25 = n \ln 0.988$
$n = -1.39/-0.012$
$n = 115.8 \sim 116$ generations