**BIOL 502 Population Genetics February 22nd, 2017**

**Spring 2017 Name:**

**Midterm Exam 1 Total: /100**

**Instructor: Arun Sethuraman**

1) Assume a large population of 1000 individuals, where we sample alleles at two independent (unlinked), diploid, biallelic genetic loci, A and B, such that the A gene has two alleles A and a, and B has two alleles B and b. The frequency of the A, a, B, b alleles are p, q, r, s respectively, such that p = 0.2, q = 0.8, r = 0.3, s = 0.7. The observed haplotype frequencies are P(AB) = 0.2, P(Ab) = 0.3, P(aB) = 0.4, p(ab) = 0.1. What would be the expected genotype frequencies of all possible genotypes after one generation of random mating? How about allele frequencies of the A, a, B and b alleles after one generation of random mating? So is this population in Hardy-Weinberg Equilibrium at these two loci (do a X2 test)? Is this population in Linkage Equilibrium at these two loci (just compute LD to determine if they are in LD/LE, you don’t have to do a test)? (7.5 +7.5 + 2.5 + 2.5 = 20 points)

2) A population of Eastern fence lizards (*Sceloporus undulatus*) at Tyson Research Center in eastern Missouri underwent a massive bottleneck in 1999 as shown below (http://press.princeton.edu/chapters/s12\_9242.pdf):

Year Count (Census size)

1997 250

1998 110

1999 26

2000 180

What are the effective population sizes in each year?

Assuming that the generation time is 1 year, is it even possible for this population to “recover” from this bottleneck (i.e. Ne = same as that in 1997) in the year 2001? Explain your answer. (10+10 = 20 points)

3) In an infinitely large population (i.e. no drift), only one type of mutation is possible at a genetic locus, such that the ‘A’ allele can mutate to the ‘a’ allele at 1x10-4 substitutions per generation. In a second population, both ‘A’ can mutate to ‘a’ at 1x10-4 substitutions per generation, and the ‘a’ allele can mutate back to the ‘A’ allele at the same rate. Assuming that the ‘A’ allele exists at frequency 1.0 in both populations, how many generations does it take each population for the ‘A’ allele to reach a frequency of 0.6, and which one is the fastest? What can you say about reversible (e.g. substitution) and irreversible (e.g. deletion) mutations in general? (20 points)

4) Why are most new mutations that arise in a population deleterious? Support and describe your answer – you’re welcome to use equations, graphs, etc. (10 points)

5) Chromosomes are known to contain several interesting features – (1) centromeric regions, where two sister chromatids are attached prior to division, (2) recombination hotspots, or regions that show elevated levels of recombination compared to the rest of the genome, (3) genomic islands of speciation, which show specific signatures of reduced migration (due to selection against migrant alleles), and (4) telomeric regions, or ends of chromosomes that are capped with telomeric repeats (for eg. TTAGGG in vertebrates). See the hypothetical chromosome below, and plot schematic levels of linkage disequilibrium relative to each other along the length of the chromosome. Explain your answer at all 4 locations (4 x 2.5 = 10 points).

\_\_\_\_\_\_\_\_\_\_\_\_\_|\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_|\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_|\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Centromere R hotspot Speciation Island Telomere

6) If a finite population of census size N = 50 were drifting at a genetic locus with three alleles, A, B and C, in frequencies p = 0.1, q = 0.3, r = 0.6 respectively, what is the probability that (a) A allele, (b) B allele, (c) C allele is lost after one generation? What can you state about the initial frequency of alleles and probability of loss? (20 points)

7) Recall that a negative Tajima’s D usually results from excess of rare variants, which can be characteristic of a population immediately after a bottleneck. This is because most common mutant variants are expected to be lost in a bottleneck. How else can common variants be lost in a population, thus resulting in a negative Tajima’s D? (Extra credit – 10 points). 

Figure from Nielsen et al. 2007, Nature Reviews Genetics, DOI: 10.1038/nrg2187, showing the values of Tajima’s D, and Θ=4Neμ estimated across the LCT (lactase) gene in humans (Asian, and European populations). This region is characterized by high homozygosity, low variability and effective population size as shown by the reduced Θ, and a skewed allele frequency spectrum (i.e. excess of rare variants) as shown by the negative Tajima’s D.

