Short Problems:

1. We survey an intergenic region (expected to be neutral) and an enzyme coding locus in two related species of fish, finding the following data:

Region	Intergenic	Enzyme Coding
Sites polymorphic within species	200	84
Sites different between species	50	78

   Note that the intergenic region and coding locus are *not* necessarily the same length.

   (a) Does this suggest selection on the coding locus, and if so, what kind? Explain briefly. (You don't need to perform a statistical test; just describe which direction the results are tending.)

   (b) Would this test still work if our interesting gene were an RNA locus that is not translated? Why/why not?

   (c) Is this test likely to be disrupted by a difference in generation time between the two species? Why/why not?

2. The human-specific locus HLA-H (previously known as HLA-AR) is similar to functional MHC loci, but has fixed mutations that seem likely to disrupt its protein structure. However, it is known to be transcribed, leading to the possibility that it still has some function.

   (a) We measure dN/dS for this gene and find that it is close to 1. Is this likely to be a functional gene or a pseudogene? Explain briefly.

   (b) We can't use between-species tests as the gene is not found in chimpanzees or other apes (it is likely a recent gene duplication of HLA-A). Can you suggest another way to determine if this locus is under selection? (Bear in mind that testing this experimentally will likely upset your IRB.)

3. Selection which periodically reverses direction–for example selecting for larger individuals in winter and smaller in summer–can maintain polymorphism, but only if it reverses often enough. Can you suggest a rule of thumb for a timescale that is definitely *too long* for a given species, so that selection will remove one set of alleles or the other rather than maintaining both?

4. Theory presented in class suggests that the chance a newly arisen allele with a heterozygote advantage s will fix is approximately $2s$, which does not depend on population size. None the less, if I have two populations of flies, one of size 10,000 and one of size 100, and I treat them with a new insecticide, the large population will become resistant more rapidly. What is missing from the theory?

Long Problem:

This problem uses the PopG simulator, found at:

http://evolution.gs.washington.edu/popgen/popg.html

You may use another simulator if you prefer, but it should be a forward simulator, not a coalescent-based backwards simulator. Please do not turn in screenshots of PopG or other simuulator runs; just tabulated data and explanations.

The approximation above gives a fixation chance of $2s$ for a newly arisen allele with a heterozygote advantage of s. We have no such handy rule for the fixation chance of a newly arisen allele with an advantage only in the homozygote–a new favorable recessive. All we know is that it should be greater than the $\frac{1}{2N}$ fixation chance for a newly arisen neutral allele and less, for a given s, than the chance for an allele that is not recessive.

1. Using the PopG simulator, check our $2s$ approximation for values of s around 0.1, using an allele which has an advantage s in both the heterozygote and the homozygote (i.e. fitnesses are 1, 1+s, 1+s). Is there a population size so low that it breaks down? (Don't forget that when you change the population size, you must change the starting frequency of the new allele so that it still exists as exactly one copy.)

2. Again using PopG, estimate the chance of fixation of a newly arisen *favorable recessive* (fitnesses are 1, 1, 1+s) for at least three values of population size and four values of selection coefficient, and tabulate your results. (I recommend at least 1000 replicate populations in each run. Population size will have to be fairly small as PopG is a forward simulator.)

3. Draw any conclusions you can from your results. Does the chance that a favorable recessive will fix depend on population size? How big does s have to be before the chance is noticably above $\frac{1}{2N}$?

4. Why would a backwards-time coalescent simulator be a bad choice for these questions?