Short Problems:

1. We have a fruit fly population with two alleles A and a at a particular locus; the allele frequency of A is 0.99. Assume we are initially in H-W and that the population size is huge. We introduce a toxin to which aa individuals are resistant, causing the fitnesses to become:

Genotype	Relative Fitness
AA	0.8
Aa	0.8
aa	1.0

   (a) (3 pts) After one generation, what is the new allele frequency of a? *0.01002. Be sure not to round away small effects like this– in a large population they matter!*

   (b) (2 pts) What is the likely ultimate fate of a, disregarding drift and mutation, if we continue poisoning our flies? *Fixation, but it will be very slow getting started.*

   (c) (3 pts) In words, why does a rare recessive allele that has a big fitness advantage increase so slowly? *It is initially present almost solely in heterozygotes, who have no advantage.*

2. Cystic fibrosis is a recessive disease that until recently was practically always lethal before adulthood. The causative locus, $CFTR$, codes for a salt channel whose absence compromises lung function and leads to fatal lung infections. Dysfunctional alleles collectively have an anomalously high allele frequency in Europeans of about 0.025, the highest of any known recessive lethal in this population.

   (a) (3 pts) If this is due to heterozygote advantage, what is the approximate advantage required? *Assume that the system is at its overdominant equilibrium. We know that equilibrium p(A) = t/(s+t). p(A) is 1-0.025=0.975. t is the disadvantage of aa and is 1, as it's lethal. Solve for s:* $0.975 = 1/(1+s)$ *so s=0.0256.*

   (b) (3 pts) Suggest as many other possibilities as you can think of for why dysfunctional $CFTR$ alleles are so frequent. *High mutation rate of the CFTR locus. Linkage of a bad CFTR allele to a favorable allele at some other locus (hitchhiking). Random drift (with so many genes to choose from, you'd expect a few to have extreme allele frequencies just by chance). Advantage of bad CFTR alleles in some subpopulation or special circumstance. I have trouble seeing how population subdivision by itself could lead to what we see.*

   (c) (3 pts) Among the dysfunctional alleles, the frequency of the deletion allele $\Delta F508$ is 68.6%, while the next most frequent allele is at 2.4%, and there is an enormous tail of rare and extremely rare variants. As far as we can tell, the severity of the disease is the same for all of the fairly common alleles (the rest are too rare to study). Does this particularly support one of your alternative ideas for the high frequency of the disease? Explain briefly. *It doesn't really support high mutation rate. It fits hitchhiking well, with the deletion allele being the one that's hitchhiking, and the rest are just random. It's consistent with random drift. Some students over-interpreted the allele frequency distribution, but we don't actually expect all equally-bad alleles to be equally-frequent except in case of overdominance. (Remember the underlying coalescent tree!)*

Long Problem:

Cavalli-Sforza and Bodmer collected the following allele frequencies for the beta-globin gene in Africans. A is the common allele worldwide. S and C cause varying degrees of sickle-cell disease in homozogyotes (and SC heterozygotes), but heterozygotes are partially resistant to malaria. Note that these are frequencies in adults; we will assume that selection has already acted on these individuals before they are sampled.

Genotype	AA	AS	SS	AC	SC	CC	Total
Observed count	25,374	5,482	67	1737	130	108	32,898
Expected count	25,615	4,967	307	1769	165	75	32,898
Fitness	0.9906	1.1037	0.2182	0.9819	0.7879	1.4400	

Calculate the relative fitness of each genotype from these data. Retain AT LEAST four digits after the decimal, because in a large population, *small fitness differences matter. Shown above. Note that these are calculated from the expectations I gave, which must have come from the original authors' knowledge of African allele frequencies as they don't quite match the data. It wouldn't be wrong to recalculate them.*

Based on these fitnesses, and assuming a huge population size so that drift can be ignored, predict:

- (3 pts) What would the equilibrium frequencies be with only A and S present? (This is the story usually told in textbooks.) *Renormalize the fitnesses so that wAS=1. This gives s=0.1025 and t=0.8023, and p(A) = 0.8867.*
- (3 pts) Can C invade a pure A population? *No. A pure A population has fitness wAA=0.9906. The initial C would all be in AC heterozygotes, which have a fitness wAC= 0.9819, so they will be selected against. The population cannot reach its maximum fitness, which would apparently be a CC genotype.*
- (3 pts) If A and S are present at their equilibrium frequency, can C invade? (Hint: When only a few C are present they will be AC and SC in proportion to the frequency of A and S. Is the weighted average of the AC and SC fitnesses better or worse than the weighted average of the population without C?) *Fitness of the A+S population at equilibrium is:*

  *wAA*pAA + wAS*pAS + wSS*pSS*

  Use Hardy-Weinberg to get the frequencies of AA, AS, and SS (at birth):

  $w \equiv 0.9906 * (0.8867)^2 + 1.1037 * 2 * 0.8867 * 0.1133 + 0.2182 * (0.1133)^2 = 1.0082$

  If the initial C are better than this, they can invade. They will be in AC and SC in proportion to the frequencies of A and S:
  *wAC * fA + wSC * fS*

  $0.9819 * 0.8867 + 0.7879 * 0.1133 = 0.9600$

  If you normalized against a different allele, or rounded differently, you will have different numbers here, but the same conclusion: it still can't invade. (Why are we seeing it in Africa, then?)

  Don't do what I did the first time, and mix renormalized fitnesses for the A and S alleles with non-renormalized fitnesses for the C alleles!
- (3 pts) Would C becoming common in the population likely lead to extinction of S, since in many ways C seems to be a better version of S? (Hint: if S could invade an $A + C$ population, it will persist in an $A + C + S$ population.)

  This problem is not ideally posed, because $A + C$ does not have a stable equilibrium. If C gets common it will simply fix; it's underdominant. Once C fixes S definitely cannot invade. But you would get a different answer if there were many A alleles still around. I accepted any reasonable quantitative answers. I didn't give full credit for guessing.

These questions do not require calculating the three-allele equilibrium, which is quite challenging. We can simply assume that if an allele yields better heterozygotes than what's already present it will be able to invade, and if it doesn't it won't. The fitness of the homozygote would be relevant for the equilibrium, but early in invasion there are so few homozygotes of the new allele that their fitness is irrelevant.