Study Guide for 2nd Exam
Anth/Bio 5221
Rogers, Seger

Gene genealogies

The questions that follow concern four copies of a neutral DNA sequence, 1000 base pairs in length. These sequences were drawn from a population of constant size, and we are interested in their gene genealogy.

1. How many coalescent intervals are in this gene genealogy?

2. What is the expected length of each interval in generations, assuming that the population size is 10,000?

3. Still assuming $N = 10,000$, what is the expected age (in generations) of the last common ancestor?

4. What is the expected total branch length of the tree? (In other words, the expected sum of the lengths of all branches.)

5. What is the expected number of segregating sites, assuming that the mutation rate is $10^{-7}$ per base pair (i.e. per site) per generation?

6. What are the expected values of $\pi$, the mean pairwise difference per site, and $\Pi$, the mean pairwise difference per sequence, using the same assumptions?

Finally, some general questions:

7. We have introduced a variety of statistics for measuring genetic variation: $H$, $\pi$, $\Pi$, and $S$. How does population size affect the expected values of these quantities?

8. Why does population size affect measures of genetic variation? (We are expecting prose here, not formulas.)

9. Theory predicts that, if one population is twice as large as another, it should have vastly more heterozygosity. Is real data consistent with this prediction? Why or why not?

10. Theory predicts that the values of $\pi$ and $S/\sum_{i=1}^{n-1}i$ should be similar, because both of these estimate the same parameter. Is this generally true in human data? Why or why not?

Mutation, drift, and population size change

11. How are the sizes and shapes of gene genealogies affected when the population size $N$ increases or decreases? (Assume selective neutrality.)

12. How are the relationships of $\pi$, $S$ and $2N$ affected when the population size has recently increased or decreased?

Site frequency spectrum and mismatch distribution

13. Be able to calculate them from data (for small, simple data sets).

14. How do they respond to changes of population size (as in the previous section)?

Selection in a large population

Each of the questions in this section might be couched in several different ways. For example, we might give you symbols for the three genotypic fitnesses and the two allele frequencies, and ask for an answer in terms of those symbols. On the other hand, we might give you numbers and ask for numbers.

15. Assuming fixed, constant fitnesses for the three genotypes at a biallelic diploid locus, what is the mean fitness of the population?

16. What are the marginal fitnesses of the two alleles?

17. What is next generation’s expected allele frequency ($p'$ or $q'$), as a function of these quantities?

18. How can an observed rate of allele-frequency change be used to estimate the relative fitnesses of two alleles (for example, $s$, if we know the value of $h$)? This was the subject of some homework problems.

Interactions of mutation, drift and selection

19. What is the fixation probability for a newly arisen neutral mutation?
20. What is the approximate fixation probability for a newly arisen mutation that is favored by a selection coefficient of $s$ (relative fitness $1 + s$) in the homozygous state, and $s/2$ in the heterozygous state? Does the population size $N$ affect our answer? If so, how?

21. What is the expected rate at which adaptive mutations will fix within a species, given a certain population size ($N$), selective advantage ($s$), and genome-wide rate of mutation ($U$) to alleles with advantages of about that size?

22. Why are most of our deleterious mutations (the ones segregating at appreciable frequencies) at least partly recessive?

23. Under what conditions will a deleterious (harmful) mutation have nearly as good a chance of fixing as a neutral mutation?

24. Do we expect more adaptive evolution in large or small populations? Explain.

25. Do we expect to find more or fewer harmful mutations segregating in a selfing plant species (for example, rice) than in a predominantly out-breeding relative? Explain.