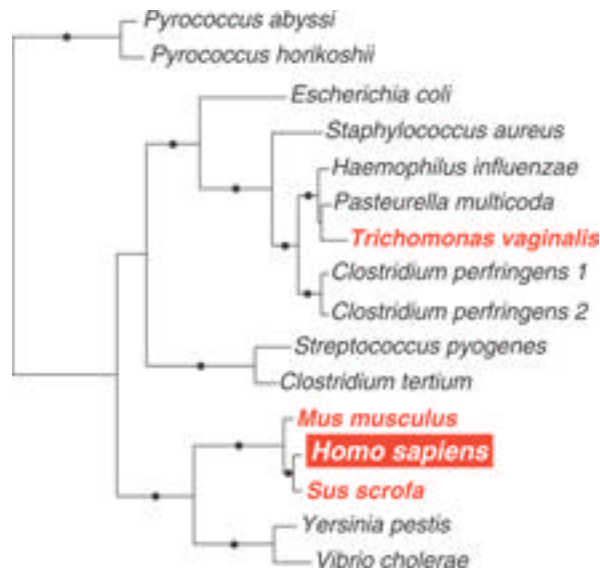


PRACTICE PROBLEMS 2

NOTE: Problems are numbered consecutively with Practice Problems 1 which included problems 1-4.

Problem 5. The figure below is a phylogenetic tree of the gene encoding N-acetylneuraminatase. All the genes are from bacteria except the four in red: *Trichomonas vaginalis* is a unicellular eukaryote parasite of vertebrates; *Sus scrofa* is the domestic pig; *Mus musculus* is the house mouse; and of course *Homo sapiens* is us. What phenomenon does this tree show?



Problem 6. Calculate the allele frequencies for the *Adh* locus in a population sample where the genotype frequencies are 0.32 *AdhF/AdhS*, 0.40 *AdhF/AdhF*, and 0.28 *AdhS/AdhS*.

Problem 7. Calculate the expected heterozygosity for each of the following loci. The data are from 120 individuals of the marine worm *Phoronopsis viralis* (Ayala et al. 1974 *Biochem. Genet.* 18:413).

| locus | frequency of allele | | | | | |
|---------------|---------------------|-------|-------|-------|-------|-------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| <i>AcpH-1</i> | 0.995 | 0.005 | | | | |
| <i>AcpH-2</i> | 0.009 | 0.066 | 0.882 | 0.014 | 0.005 | 0.024 |
| <i>Est-5</i> | 0.483 | 0.396 | 0.122 | | | |

Problem 8. Herbert did starch gel electrophoresis of the plastid elongation factor Tu, Marisa amplified and sequenced the coding region of the *tufA* gene from each of five

clones of the alga *Polytoma obtusum*. The genes were identical except in one region, for which the sequences are shown below. These sequences contain only complete codons.

```
1 AAAACGACATTAACAGCAGCAATAACTATGACTTTAGCAGCACGTGGAAACTCCGTAGGTAAAAAATATGAAGAA
2 AAAACGACATTAACAACAGCAATAACTATGACTTTAGCAGCACGTGGAAACTCCGTAGGTAAAAAATATGAAGAA
3 AAAACGACATTAACATCAGCAATAACTATGACTTTAGCAGCACGTGGAAACTCCGTAGGTAAAAAATATGAAGAA
4 AAAACGACATTAACAGCAGCAATAACTATGACTTTAGCAGCACGTGGGAACTCCGTAGGTAAAAAATATGAAGAC
5 AAAACGACATTAACAGCAGCAATAACTATGACTTTAGCAGCACGTGGAAACTCCGTAGGTAAAAAATATGAAGAC
```

(a) How many alleles of the gene were identified by sequencing? Call the alleles *tufA1*, *tufA2*, etc. and indicate which clones have which allele.

(b) How many alleles would Marisa have found if, instead of sequencing, she had done restriction analysis of the gene with the enzyme *BbvI* which cuts the site GCAGC? Call the alleles *tufA1*, *tufA2*, etc. and indicate which clones have which allele.

(c) How many alleles would Marisa have found if, instead of sequencing, she had done electrophoresis of elongation factor Tu, the protein encoded by *tufA*? Assume she did the electrophoresis two ways: first, simple electrophoresis separating by charge alone; second, electrophoresis under enough different conditions to detect all amino acid differences.

Problem 9. Consider a diploid population of size $N = 50,000$.

(a) What is the fixation probability of a neutral mutation in this population?

(b) Suppose the mutant allele is not lost immediately but its frequency increases by drift to 0.1. Now what is its fixation probability?

(c) Suppose the frequency of this same allele drifts to 0.8. Now what is its fixation probability?

(d) What would be the answers for a haploid population?

Problem 10. Fixation and coalescent times.

(a) Calculate the mean time to fixation for a new mutation in diploid populations of size 10^3 , 10^5 , and 10^6 .

(b) What do you think is the mean fixation time for a new mutation in a population of 10^8 *Chlamydomonas reinhardtii* cells? (This alga is haploid.)

(c) What is the mean time to the coalescent of two copies of a gene in the *Chlamydomonas* population?

Problem 11. How much heterozygosity would you expect to see under the neutral hypothesis, if you looked for electrophoretic variants for β -globin in a large sample of rabbits from a population, assuming that the rabbit population is in mutation-drift equilibrium? Warning: don't use the approximate version of the equation! Assume that the mutation rate is 5×10^{-5} mutations per gene per generation and the effective population size is

(a) 10^3

(b) 10^5

(c) Assuming that this mutation rate is approximately correct for electrophoretic variants of β -globin, and that β -globin is a fairly typical protein, which of the preceding effective population sizes do you think is closer to correct for this population of rabbits?

(d) The formula for the expected value of the nucleotide diversity at mutation-drift equilibrium is the same as the equation for expected heterozygosity, except that the

mutation rate is mutations per base pair per generation. This value is about 5×10^{-9} for vertebrates. Using this rate, suppose two different geneticists clone and sequence a globin gene taken from a rabbit population. When they compare their sequences, what proportion of base pairs do you expect to be different in their sequences if the effective population size is 10^5 for the rabbit population?

(e) Consider the first base pair in the β -globin gene: what is the probability that it is different in the two rabbit genes?

Problem 12. Assume that a haploid species has a population size of $N = N_e = 10^3$; that the total mutation rate is 5×10^{-9} per bp per year per gamete; that the ratio of detrimental to advantageous mutations is 9:1; and that the mean selection coefficients of detrimental and advantageous mutations are -0.001 and $+0.001$, respectively. Calculate the overall rate of molecular evolution for two cases: (a) 20% of all mutations are neutral; and (b) 80% of all mutations are neutral.

Problem 13. Li and collaborators calculated the average sequence divergence for synonymous substitutions in a large number of protein-coding genes, and for a large number of comparisons of mammals of different orders. The average divergence, corrected for multiple hits, was 0.744. All the orders of mammals are believed to have had a common ancestor about 80 My ago. Calculate the average rate of synonymous substitution for mammalian genes. Then estimate the mutation rate for mammals.

Problem 14. In mammals the rates of synonymous substitution differ between nuclear and mitochondrial genes. The rate is 4.6×10^{-9} bp substitutions per bp per year in the nucleus, and about 50×10^{-9} in the mitochondrion. What, if anything, can you infer from these data about the mutation rates in mammalian nuclei and mitochondria?

Problem 15. Assume that the molecular clock ticks at a rate of 5×10^{-9} bp substitutions per bp per year. On a volcanic island you find two species of *Drosophila*, descended from one species that colonized the island some time after it first rose out of the ocean. You sequence the *Adh* genes of the two species and find they show 5 synonymous substitutions in 1 kbp.

- (a) What is the frequency of synonymous substitutions between the two species?
- (b) How long ago did the two species diverge?
- (c) How old is the island?