

Problem Set 4, Bio 4181: Due Oct. 13, 2009

1. A sample of 20 homologous genes from a tetraploid species is drawn at random from a large population with variance effective size of 1,000 and inbreeding effective size of 500. Assume all individuals are self-compatible hermaphrodites.

a). What is the probability that two of the sampled genes coalesced in the previous generation?

Prob.(coalescence in the previous gen.) = $\frac{n(n-1)}{2xN}$ where $x=4$ (1 pt), $n=20$, and $N=500$ (only the inbreeding size is relevant, 1 pt), so Prob.(coal. prev. gen.)=0.095 (1 pt)

b). What is the probability that the first coalescent event occurs 4 generations ago?

From Box 5.1, Prob.(first coalescence in t generations) = $\left(1 - \frac{n(n-1)}{2xN}\right)^{t-1} \frac{n(n-1)}{2xN}$ (1 pt) where $x=4$, $n=20$, and $N=500$ to yield 0.0704 (2 pts)

c). What is the expected time (in generations) between the second to last and ultimate coalescent event?

From Box 5.1, the expected time between the second to last and last coalescent event is $xN=2000$ generations (2 pts)

d). What is the expected time to coalescence for a random pair of these genes?

This is also $xN = 2000$ generations (2 pts).

e). Chloroplast DNA is inherited as an effective haploid element in this species. Given a random sample of 20 chloroplast DNA molecules, redo calculations a) through d).

All is the same except $x=1$ (1 pt), so a) 0.38 (1 pt), b) 0.0906 (1 pt), c) 500 (1 pt), d) 500 (1 pt)

2. Two X-linked genes are randomly drawn from a population of variance effective size 500 and inbreeding effective size 1000. The neutral mutation rate at this locus is 10^{-5} per DNA replication. What is the probability that the two sampled genes are identical by descent, assuming that only neutral mutations occur?

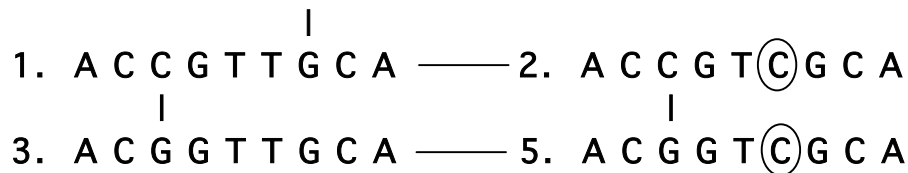
Equation 5.14 gives the probability of non-identity by descent (1 pt), with $x = 1.5$ and $N_{ef} = 1000$ (1 pt), so $\theta = 0.03$ (1 pt), and prob. of non-ibd = 0.029 (1 pt). Hence, the prob. of ibd = $1 - 0.029 = 0.971$.

3. A DNA region of 9 nucleotides is sequenced, with variable sites being found that define 5 haplotypes as follows:

1. A C C G T T G C A
2. A C C G T C G C A
3. A C G G T T G C A
4. A C C G T T A C A
5. A C G G T C G C A

- Construct a haplotype tree(s) assuming the least number of mutational events (maximum parsimony) and no recombination.
- Is there any ambiguity in this tree, and if so, how many distinct tree topologies are there under maximum parsimony?
- Is the tree consistent with the infinite sites model? Justify your answer.
- Is there any homoplasy in the maximum parsimony tree(s)? Justify your answer.
- What is the Jukes-Cantor genetic distance between haplotypes 4 and 5?

4. A C C G T T A C A



a. The topology is given above, with 2 differing from 1 at the circled site, and 5 differing from 3 at the circled site. (4 pts)

b. There is ambiguity because 1, 2, 3 and 5 form a loop (1 pt). Under maximum parsimony, there are four ways of breaking this loop (break 1-3, 1-2, 3-5 or 2-5), so there are 4 maximum parsimony trees (2 pts).

c. Either the 3rd or the 6th site has at least 2 mutational hits, so the infinite sites model is violated (1 pt).

d. Both of the possible multiple mutations are to the same state, so there is homoplasy in this tree (2 pts.)

e. $\pi=3/9= 1/3$ (1 pt)., $D_{JC} = -\frac{3}{4} \ln(1 - \frac{4}{3} \pi)$, so $D_{JC}=0.44084$ (2 pts). This is the number of expected mutations per nucleotide, so if you wanted the number of expected mutations for the entire 9 bp region, multiple by 9 to get 3.97. Either answer is acceptable.

4. Two infinite-sized demes both have the A and a alleles at an autosomal locus, but with the frequency of A being 0.2 in population 1 and 0.7 in population 2. Suppose that now the two populations begin to symmetrically exchange gametes at a rate of 10% per generation.

a) What are the expected allele frequencies in demes 1 and 2 after two generations of gene flow?

From the book, $p_1' = (1-m)p_1 + mp_2$ and $p_2' = (1-m)p_2 + mp_1$. Hence, after 1 generation $p_1(t=1)=.25$ and $p_2(t=1)=.65$ (1 pt). Use iteratively for next generation to get $p_1(t=2)=.29$ and $p_2(t=2)=.61$ (2 pts).

b) What is the equilibrium frequency of A in demes 1 and 2?

At equilibrium $p_1=p_2=[p_1(t=0)+p_2(t=0)]/2=0.45$ (2 pts).

5. Two copies of homologous genes from an autosomal locus are sampled at random from the same deme of inbreeding effective size 40 and variance effective size 50. The deme itself is just one deme of many that fit an island-model of subdivision with $m = .01$.

a). What is the probability that these two genes coalesce before either lineage experienced gene flow?

Use equation 6.11 (1 pt) to first calculate the prob. of gene flow before coalescence as 0.615 (1 pt), so the prob. of coal before gene flow is 0.385 (1 pt).

b). What is the ratio of expected coalescence times of a pair of genes sampled from the same local deme to the expected coalescence time of a pair of genes sampled at random from the total population?

Use equation 6.13 to obtain the ratio of 0.615 (2 pts).

6. Five populations are surveyed for genetic variation at an autosomal locus with the following results:

	AA	Aa	aa
Population 1	56	288	156
Population 2	156	288	56
Population 3	200	600	200
Population 4	608	384	8
Population 5	16	768	1216

What are f_{is} , f_{st} , and f_{it} ?

Measure f_{is} as deviation from HW genotype frequencies within each population. In each case, get $f_{is} = -0.2$ (2 pts).

	AA	Aa	aa
sums:	1036	2328	1636
freq.	0.2072	0.4656	0.3272

Measure f_{it} as deviation from HW in pooled population (above). In the pooled population, the frequency of the A allele is 0.44 (1 pt). Get $f_{it} = 0.055$ (2 pts).

Since $(1-f_{it}) = (1-f_{st})(1-f_{is})$, solve for $f_{st} = 1-(1-f_{it})/(1-f_{is}) = 0.213$ (2 pts). [Alternatively, can calculate weighted variance of p over populations and divide by $\bar{p}\bar{q}$]

7. A species has a 1-dimensional stepping stone model of population structure with local inbreeding effective sizes of 100 and variance effective sizes of 50. 20% of the gametes are exchanged only among adjacent demes, and 0.1% are exchanged at random over all local demes.

a) What is the equilibrium f_{st} ?

Use equation 6.34 with $N_{ev} = 50$, $m_l=0.2$, and $m_{\infty}=0.001$ to obtain $f_{st}=0.2$ (2 pts).

b) Now suppose that the local exchange rate of gametes is doubled to 40%. What is the equilibrium f_{st} ?

Use equation 6.34 with $N_{ev} = 50$, $m_l=0.4$, and $m_{\infty}=0.001$ to obtain $f_{st}=0.15$ (2 pts).

c) Now suppose that the local exchange rate of gametes is restored to 20% but the long-distance exchange rate is doubled to 0.2%. What is the equilibrium f_{st} ?

Use equation 6.34 with $N_{ev} = 50$, $m_l = 0.2$, and $m_\infty = 0.002$ to obtain $f_{st} = 0.15$ (2 pts).