**BIOL 502 Population Genetics**

**Spring 2017**

**Midterm 2 Review**

**Selection**

1. For each of the following fitness sets plot allele frequency versus time for 100 generations, starting at an allele frequency of 0.5. What would be the expected equilibrium allele frequencies in each case? Compute these, and explain them in biological terms.

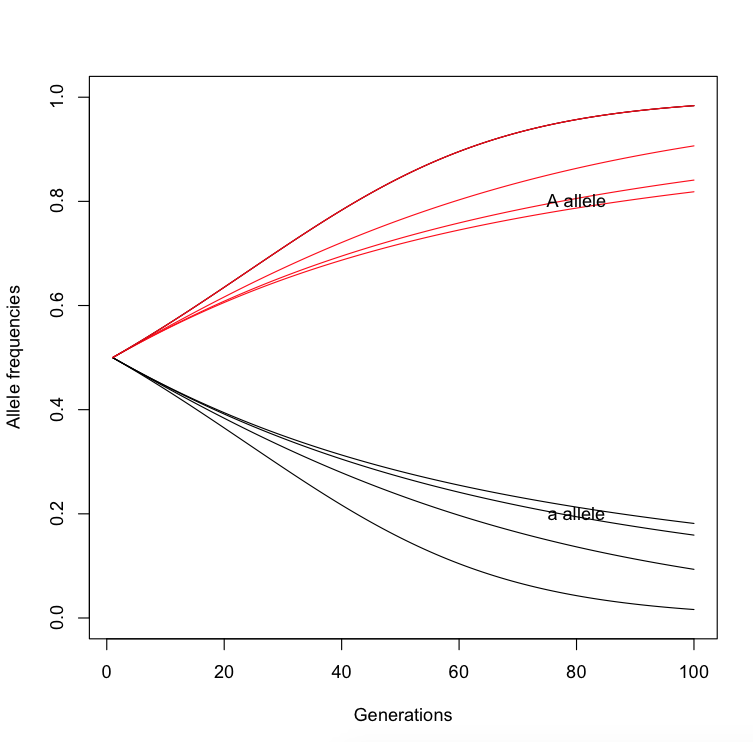
w11 w12 w22

1 1 0.95

1 0.995 0.95

1 0.98 0.95

1 0.95 0.95

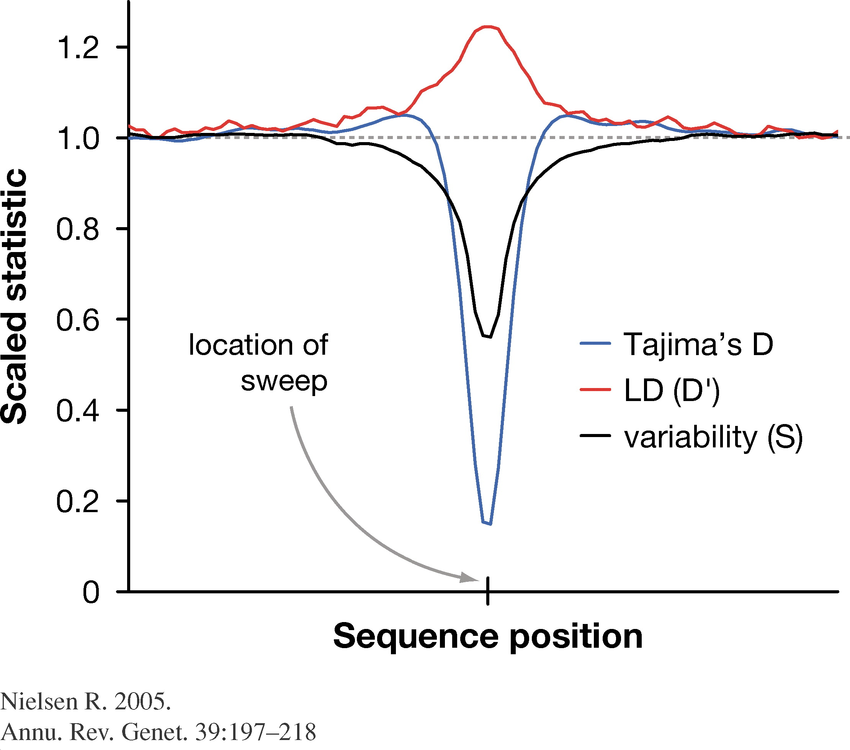


As mean fitness of heterozygotes decreases, relative to the homozygote dominant genotype, the time to equilibrium (or fixation) decreases, i.e. the allele reaches equilibrium faster.

2. If a population has a mutation rate of u = 10^-6 mutations per base generation, and a selection coefficient of s = 0.4, what would be the expected equilibrium frequency of the recessive allele (if it’s completely recessive)?

Equilibrium frequency = sqrt(mu/s) = 0.001581139

3. What is the probability of fixation of a favorable mutation with s = 0.1, when N = 100, Ne = 20, 100, and 200? Interpret your results.



If N = 100, Pr(fixation) = 0.1812692

If N = 20, Pr(fixation) = 0.1813301

If N = 200, Pr(fixation) = 0.1812692

If N = 2, on the other hand, Pr(fixation) = 0.3291788

In other words, as population size increases, the effect of selection is decreased.

4. If you were to compute the dN/dS ratio in windows across the genome in a human population (say Central European), and plot it (X axis = dN/dS, Y axis = # of protein coding regions), knowing what you’ve learned about drift, selection, and mutations, what would this plot be expected to look like?

5. What is going on at the genomic position marked by the arrow? Explain your answer with respect to Tajima’s D, LD, and Variability (differentiation, say measured as Fst).

Low Tajima’s D, indicates an excess of rare variants compared – this could potentially indicate signatures of recent bottleneck, or a selective sweep.

Low Fst indicates that this region is less differentiated than the remainder of the genome. This could indicate one of several things – (a) selective sweep at this locus, driving effective population size and genomic diversity down, (b) negative, or background selection at this locus, fixing variation around it, (c) excessive migration at this locus (i.e. positive selection for migrant alleles).

High LD can indicate several things: (a) non-random association of alleles around this locus, indicating reduced recombination rate, (b) excessive migration at this locus, (c) selection for the heterozygote at this locus, which can artificially increase frequency of heterozygotes, relative to the allele frequencies.

**Inbreeding**

1. How does the inbreeding coefficient vary in a small population versus a large population? Explain showing equations, and interpret it biologically.

../../../../../Desktop/Screen%20Shot%202017-04-10%20at%209.52.21%20

So as Ne increases, the inbreeding coefficient decreases.



2. What is the inbreeding coefficient of the offspring of a full-sibling mating? If this individual also mates with its sibling, what would be the inbreeding coefficient of their offspring? What can you state about the rate of change in inbreeding coefficient due to regular full-sibling mating?

A B

| X |

C D

X

E

Assuming that A and B are not inbred, there are n = 2 paths (CAD,CBD). So F = 2 x (1/2)3 = 1/4.

If you continue this for one more generation,

A B

| X |

C D

| X |

E F

X

G

Now there are many more paths – ECF, EDF, ECADF, ECBDF. So F in this generation = 1/4 + (1/2)5 + (1/2)5 = 5/16.

Inbreeding coefficient increases with more inbreeding.

3. Compute (a) coefficient of consanguinity, (b) coefficient of relatedness between a grandparent and a grandchild.

A

|

B

|

C

Consanguinity = sum over 4 alleles (1/2)(1)(1/2)(1/4) = ¼

Relatedness = (1/2)^2 = ¼

4. What is the difference between inbreeding and subpopulation structure? Explain.

Inbreeding describes mating between related individuals, whereas subpopulation structure occurs due to non-random mating of individuals in the same geographical area (related or unrelated).

**Population Structure and Migration**

1. Does the Wahlund Effect change with increasing degree of subpopulation structure within a total population? If not, why? If so, why?

No – Wahlund effect doesn’t change, since it depends on the mean allele frequency across the total population, computed as sum (all K populations) (wk pk), where wk = nk/N is the proportion of individuals in the total population that belong to subpopulation k, and pk is the frequency of the allele in the kth subpopulation. So the Wahlund effect depends on the frequencies, and the relative proportions of individuals within each population, and not on the degree of population structure (i.e. K) itself.

2. In an island model, if you have a total of three islands, exchanging genes at an equal rate m with each other, but you only sample individuals from two of the islands (the third unsampled population is often called a ghost population), how would you miscalculate the expected allele frequency after one generation? Show this mathematically.

Let’s consider that there are three islands, exchanging genes at the rate of m in every generation. Let frequency of alleles in each island be pt (in generation t). So if you consider any one single population, the frequency of this allele in the next generation = mp(total) + pt(1-m), where p(total) = the frequency of the allele across all three islands = (p1+p2+p3)/3. If you didn’t sample the third population, this would be miscalculated as (p1+p2)/2, hence incorrectly computing the allele frequency in any one sampled island after one generation.

3. In a plot of isolation by distance, instead of observing a positive slope (i.e. increased isolation with increased geographical distance), you observe a negative slope. How is this biologically possible? Explain your answer.

If there’s a negative slope, subpopulations that are close to each other are more differentiated compared to subpopulations that are far away from each other. This could indicate ancestral polymorphisms segregating at distant subpopulations at higher frequencies, also known as incomplete lineage sorting. Or it could indicate migration over long distances.

4. Interpret this plot of genetic population structure:



5. If a population is receiving migrants at the rate of m per generation, and mutating at the rate of u mutations per generation per base, what would be the probability of IBD between two randomly sampled alleles in this population after one generation? Derive this.