The format of final will be like that of the other exams. It will be longer, however, since you have more time. It is cumulative, but we will weight recent material more heavily. This study guide does not cover material from exams 1 and 2. (Use the old study guides.) If you have questions, feel free to email any of us.

Multiple loci

1. Be familiar with Gillespie’s notation \((x_1, x_2, x_3,\) and \(x_4)\) for frequencies of two-locus haplotypes, where each locus has two alleles.

2. Be able to re-express these in terms of the \(p_A, p_B,\) and \(D.\)

3. Under random mating, what are the frequencies of all possible two-locus genotypes \((AB/AB,\) \(AB/aB,\) and so on)? You should be able to answer this using either of the two notations: either in terms of \(x_1, x_2,\) \(x_3,\) and \(x_4,\) or in terms of \(p_A, p_B,\) and \(D.\)

4. If we gave you the frequencies of these four gamete types, you should be able to give us the allele frequencies and \(D (the coefficient of linkage disequilibrium). Or the other way around.

5. What are the relationships among these two measures of linkage disequilibrium: \(D\) and \(r_H?\) (Here, \(r_H\) refers to the gametic correlation between two loci. We used this symbol in the lab manual. Gillespie uses \("r\" for the same purpose.)

Suppose that in an initial generation, \(x_1, x_2, x_3,\) and \(x_4,\) are equal to 0.2, 0.3, 0.4, and 0.1 and that the recombination rate is \(c = 1/10.\)

6. What are \(p_A, p_B,\) and \(D\) in this initial generation?

7. What are the expected values of these quantities after 1 generation? After 100 generations?

8. In lecture, Rogers explained that (a) selection does not generate linkage disequilibrium (LD), but nonetheless (b) LD is useful in detecting ongoing selective sweeps. Why is this not a contradiction? (See “Why linkage disequilibrium helps us find selective sweeps” on the class website.)

9. Locus \(B\) has two alleles: \(B_1\) (with frequency \(p_1),\) and \(B_2\) (with frequency \(1-p_1\)). A new mutation arises at a linked locus, \(A.\) What is the probability that this mutant chromosome carries allele \(B_1?\)

10. Mammalian genomes (including ours) are typically about 3 Gbp (billion base pairs) in length, but most of that DNA is non-functional “junk” that doesn’t code for proteins or structural RNAs, and doesn’t play any important role in the regulation of gene expression. Some of these neutral nucleotides are close to functional genes, and others are far away. How would you expect proximity to functional genes to affect the variability of these neutral nucleotides? Are the ones near genes more likely to be polymorphic, or less? Explain your reasoning.

11. In Gillespie’s standard setup, there are four gamete types \(A_1B_1, A_1B_2, A_2B_1,\) and \(A_2B_2,\) with frequencies \(x_1, x_2, x_3,\) and \(x_4.\) Suppose that, in generation 1, half the gametes are \(A_1B_1\) and half are \(A_2B_2.\)

(a) What is \(D?\)

(b) Now assume that selection operates at the gamete stage, that alleles \(B_1\) and \(B_2\) are neutral, that allele \(A_1\) has fitness \(1+s\) relative to \(A_2,\) and that the recombination rate is \(c = 1/3.\) What are the expected values of the four gamete frequencies in the following generation? (Hint: Use the formula for selection at the gamete stage, which we gave in lecture.)

12. When an allele sweeps to fixation at locus \(A,\)

(a) what happens to the frequencies of neutral alleles at linked loci? (b) What happens to \(D?\)

13. Linkage disequilibrium is often obvious in tables of DNA sequence (or SNP) data. For examples, see our slides. Be able to recognize it when you see it.

14. According to recent studies, several human loci (at least) are currently undergoing selective sweeps. Be prepared to discuss their evidence and its implications.
15. What are lactose, lactase, lactose intolerance, and lactase persistence? Be familiar with the evidence (archaeological and genetic) suggesting that lactase persistence has been strongly favored by selection in the recent history of Europeans.

16. What is genetic draft? What problem was it invented to explain? How is it affected by the following parameters: $N$, $\mu$, $c$ (recombination rate)?

**Inbreeding**

17. What is inbreeding depression, and what causes it?

18. The inbreeding coefficient ($F$) can be interpreted as a measure of either (a) departure from Hardy-Weinberg genotype frequencies, (b) inbreeding within a pedigree, (c) inbreeding between random individuals within a population, or (d) the effect of genetic drift. Be familiar with all of these uses.

19. Be able to express the frequencies of the three genotypes at a biallelic locus ($P_{AA}$, $P_{Aa}$, and $P_{aa}$) in terms of $p$ and $F$. Know what these quantities represent.

20. The formulas in the preceding question can be used to represent the effect of non-random mating in a single generation, or the effect of drift across many generations. How do the interpretations of $p$ and $F$ change in these two situations? (Hint: your answer should involve the “reference generation.”)

21. Genotype frequencies can be written in two algebraically equivalent forms. For example, $P_{AA}$ can be written either as $p^2 + p(1 - p)F$ or as $Fp + (1 - F)p^2$. Be familiar with both forms.

22. What is the connection between inbreeding and genetic drift?

23. Consider the following genealogy:

   
   |   F   | G   |
   | \  / |
   | \ \ |
   | \ \ |
   | / \ |
   |   D  |
   |   E  |
   |   |
   |   |
   |   |
   | B   |
   |   C |
   | \  / |
   | \  / |
   |   A |

   In this genealogy, A is the offspring of B and C, which are the offspring of D and E, and so on. (a) The two genes that united to form A may be identical by descent (IBD) from which ancestors? (b) For each of these ancestors, calculate the probability of IBD. (c) What is A’s inbreeding coefficient? (d) What is the coefficient of kinship of D and C? (e) In what way, if any, do we expect A to differ genetically from typical members of the population? Explain briefly. (f) In what way(s) might we expect A to differ phenotypically from typical members of the population? Explain briefly. Answers: (a) F, and G; (b) 1/32 for F or G; (c) $F = 2 \times 1/32 = 1/16$; (d) $f_{DC} = 1/8$. (e) Individual A will be heterozygous at fewer loci than will the average member of the population, because there is an appreciable probability that the gene copies he inherited from his two parents were identical by descent from their common ancestor, individual F. (f) Individual A is likely to be somewhat shorter than the average individual and somewhat less healthy, because he will have fewer loci that exhibit heterozygote advantage and more that are homozygous for recessive deleterious alleles.

24. Figure 1 shows the genealogy of Francis Darwin, one of Charles Darwin’s sons. What is his inbreeding coefficient?

**Non-random mating**

25. Suppose that a population consists of several subdivisions (or groups) of equal size, within each of which mating is at random. The mean
29. Among continental human populations, estimates of $F_{ST}$ are usually close to 1/9. Assume that this represents an equilibrium between migration and genetic drift under the Island Model of population structure. What does this imply about the number $Nm$ of migrants between pairs of populations in each generation?

**Modern human origins**

30. What are the multiregional and replacement hypotheses? What evidence supports each hypothesis? What evidence weakens each?

31. How have the Neanderthal and Denisovan genomes changed the debate about modern human origins?

**Quantitative traits**

32. What are quantitative traits? Why are they often normally (or near-normally) distributed? What is the phenotypic variance ($V_p$)? What are the relationships (mathematical and biological) among the variance components $V_p$, $V_g$, $V_a$, $V_d$ and $V_e$? What is the narrow-sense heritability ($h^2$) and why is it important? How can each of the variance components be estimated for real populations? Why do estimates for one population have little validity for another population of the same species? (i.e., why do we have to make the estimates anew, from actual data, to be confident of their values in a given population?)

33. Related question: Why does the fact that a trait is highly heritable within each of two populations

---

**Figure 1: Genealogy of Francis Darwin**

- Robert Waring Darwin
  - Charles R. Darwin
  - Josiah Wedgwood
  - Susannah Wedgwood
  - Francis Darwin
    - Josiah Wedgwood, II
    - Emma Wedgewood
    - Sarah Wedgewood
  - Elizabeth Allen

- Emma Wedgewood
  - Sarah Wedgewood

---

of group allele frequencies is 1/2 and their variance is 1/4. What is the expected heterozygosity in the population as a whole? (Hint: See the Wahlund Effect, discussed in lecture and in the text.)

26. What is $F_{ST}$, and what does it measure? How does it increase with time among a group of isolated populations, such as those in Buri’s experiment? What equilibrium does it reach if the effects of drift are opposed by gene flow? Be able to calculate it from data.

27. Suppose that we ran an experiment like Buri’s—one in which genetic drift was the only force at work, the effective population size was constant at $N = 50$ within each group, and our organisms were diploid. Unlike Buri’s organisms, our mated at random. The initial heterozygosity was $H_0 = 1/2$, and we ran the experiment for 5 generations.

   (a) What is the expected heterozygosity in the final (the 5th) generation?
   (b) What is the expected value of $F_{ST}$?
   (c) What is the average inbreeding coefficient of individuals in the 5th generation relative to the initial (the 0th) generation?

28. Use the data in the table below to estimate $F_{ST}$. (Hint: you only need the last line.)

<table>
<thead>
<tr>
<th>Population</th>
<th>$p$</th>
<th>$p^2 + q^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>African</td>
<td>0.734</td>
<td>0.610</td>
</tr>
<tr>
<td>European</td>
<td>0.496</td>
<td>0.500</td>
</tr>
<tr>
<td>Australian</td>
<td>0.330</td>
<td>0.558</td>
</tr>
<tr>
<td>East Asian</td>
<td>0.259</td>
<td>0.616</td>
</tr>
<tr>
<td>Average</td>
<td>0.455</td>
<td>0.571</td>
</tr>
</tbody>
</table>
NOT allow us to conclude that an observed average difference between the two populations is “caused by genes?”

34. Response to selection can be predicted either from

\[ R = h^2 S, \quad \text{or from} \]
\[ \Delta x = V_a \beta \]

Gillespie discusses only the first one, but both are covered in Seger’s notes on quantitative traits (see class website). Be familiar with both formulations and able to manipulate them.

35. What evidence suggests that long-term rates of quantitative trait evolution are not limited by amounts of heritable genetic variation?