

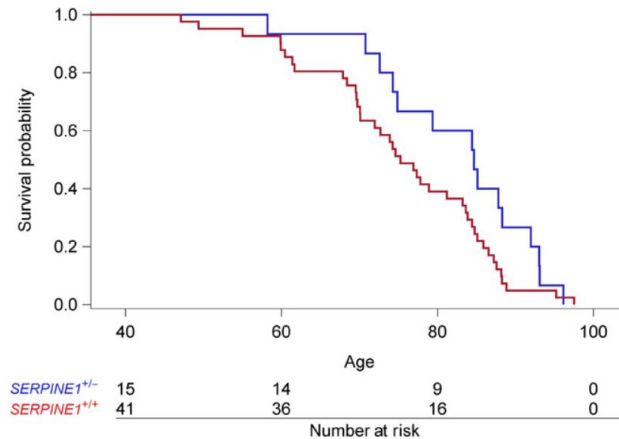
You've elected to have a full week to work on this assignment collaboratively, and with all the resources of the internet at your disposal. This includes everything short of cheating (googling, reading the referenced papers, getting background information by talking to me, discussing with one another, etc.)

The answer to each question needs to convey your understanding as fully as you can. This is especially true for the questions with word limits. A vague description gets a poor grade.

No new R coding is *required* for any of these questions, however you may find that some of them are significantly easier to answer using some minor modifications of R code you've already written or that I give you (rather than trying to work them out by hand).

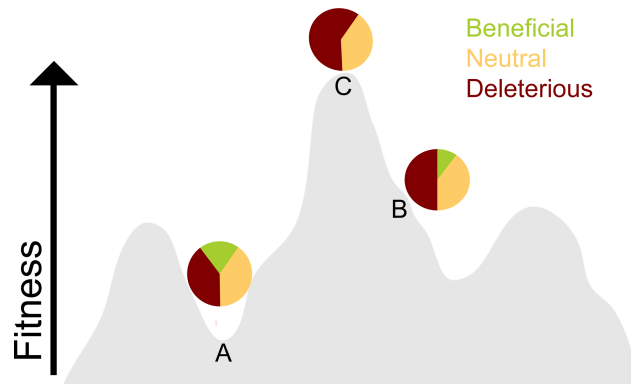
**You must each turn in your own assignment.** You can work together to discuss potential answers, but the answers you turn in to me must be in your own words. Plagiarism/copying is the only way to cheat, any resource is fair game, any amount of working together is also fair. Just don't directly copy.

Good luck!

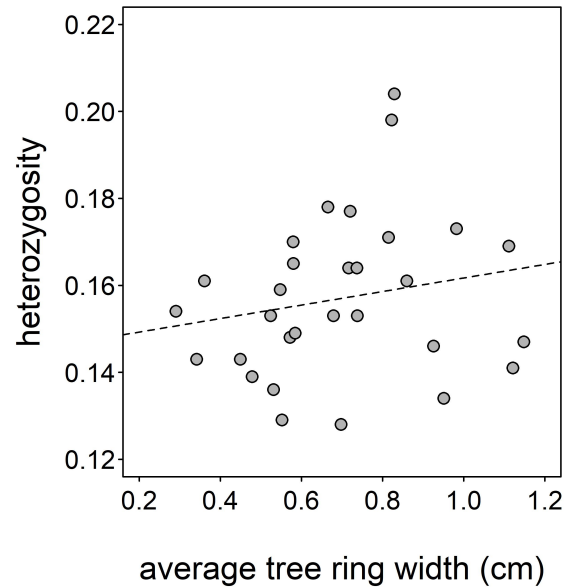
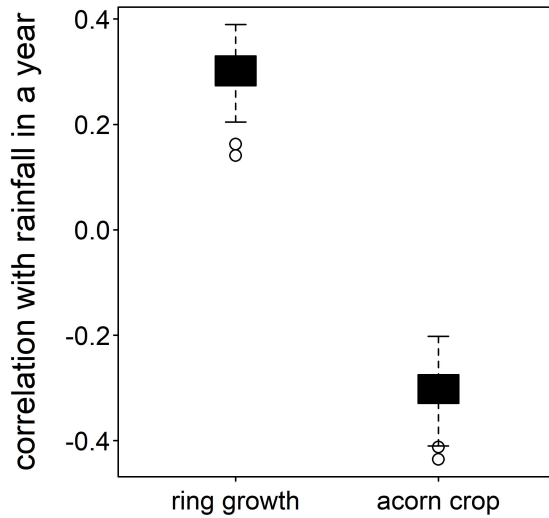


1. (4 points) One speaker described a recently discovered mutation called *SERPINE1* in the Amish community of southern Indiana which dramatically increases longevity (see graph above). The study that discovered this mutation genotyped 177 randomly sampled individuals and found that, at this locus, 127 were homozygous for the major allele (*aa*), 43 were heterozygotes (*ab*), and 7 were homozygous for *SERPINE1* (the minor allele; *bb*). All of the *bb* individuals were under age 40 and had no children. The *SERPINE1* allele is a mutated form of the 402 amino acids long PAI-1 protein. A 2020 study by Michael Kessler et al. estimated a mutation rate of  $1.13 \times 10^{-8}$  per base pair in the Amish. Dr. Nekeisha Nixon in 2013 found that Amish women have an average of  $7.7$  children, with an average age at birth of  $30.5 \pm 6.7$  years old. There are approximately 275,000 Amish people currently living in the US and Canada. The  $F$  value for Amish populations was estimated to be about 0.017 Cabellero in 1994 and to be about 0.035 by Van Hout in 2011 a generation later. The change in  $F$  in a generation is  $\frac{1}{2N_e}$ . A 2015 study by Braxton Mitchell and colleagues showed that, based on 1000 loci,  $\approx 82\%$  of alleles present in the founding Amish generation have been lost over the course of the past 13 generations. We need to make some simple assumptions in order to proceed. None of these are true, but they're close enough to true that you can get a reasonable answer using them. Assume all numbers given above are precise and accurate. Assume maternal age at birth follows a perfectly normal distribution (so that you can use `dnorm()` in R to compute the percentage of births past any specific age). Assume that all genotypes are identical in every way before age 40, and that the only direct effect is the longevity effect shown here. Assume the chance a child dies before reaching reproductive age is 0.5% for those born to mothers under 40, and 2% for those born to mothers over 40. Assume this population is perfectly monogamous, and that men and women only marry if they are the exact same age. Assume this population is perfectly insular, with no individuals ever leaving or new individuals ever joining. Assume there are no other alleles for this locus. Using the assumptions and data above:

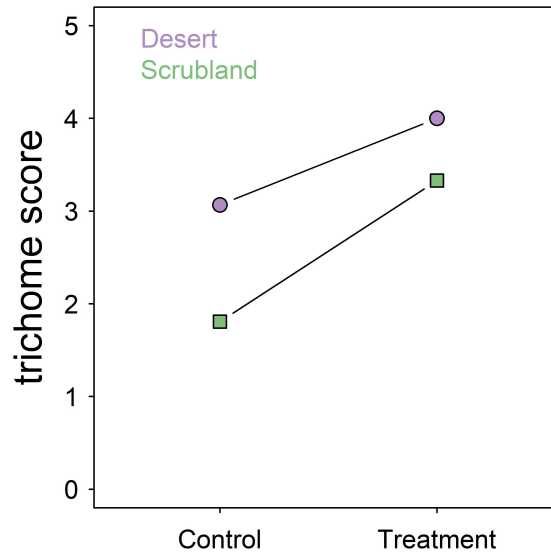
- Determine the frequency of allele *b* (*SERPINE1*) in the population and explain how you did so.
- Determine the fitnesses relative to the *aa* of each genotype and explain how you did so.
- Determine the  $s$  and  $h$  coefficients for allele *b* (*SERPINE1*) and explain how you did so.
- Determine the effective population size and explain how you did so.
- Assuming a generation is 21 years, determine approximately how long this locus will be polymorphic, and what the most likely allele to go to fixation will be (assume no mutations).



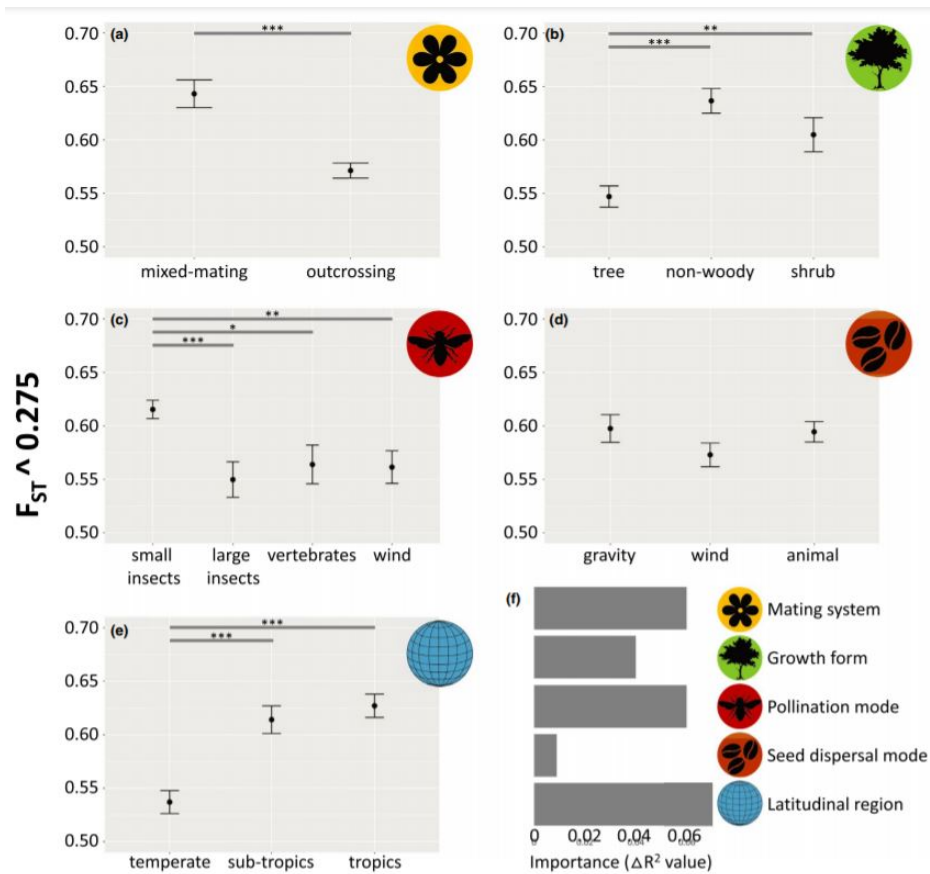
**2.** (1 point) *One speaker gave an excellent review of basic population genetics forces. Above is a figure from a paper published in PLoS in 2018 by Dr. Siobain Duffy. The y-axis shows fitness, the x-axis shows genotype, and the pie charts show the relative frequency of different  $s$  values for mutations. In 100 words or fewer, explain why the proportions of mutation types differ between A/B/C as clearly and precisely as you can.*



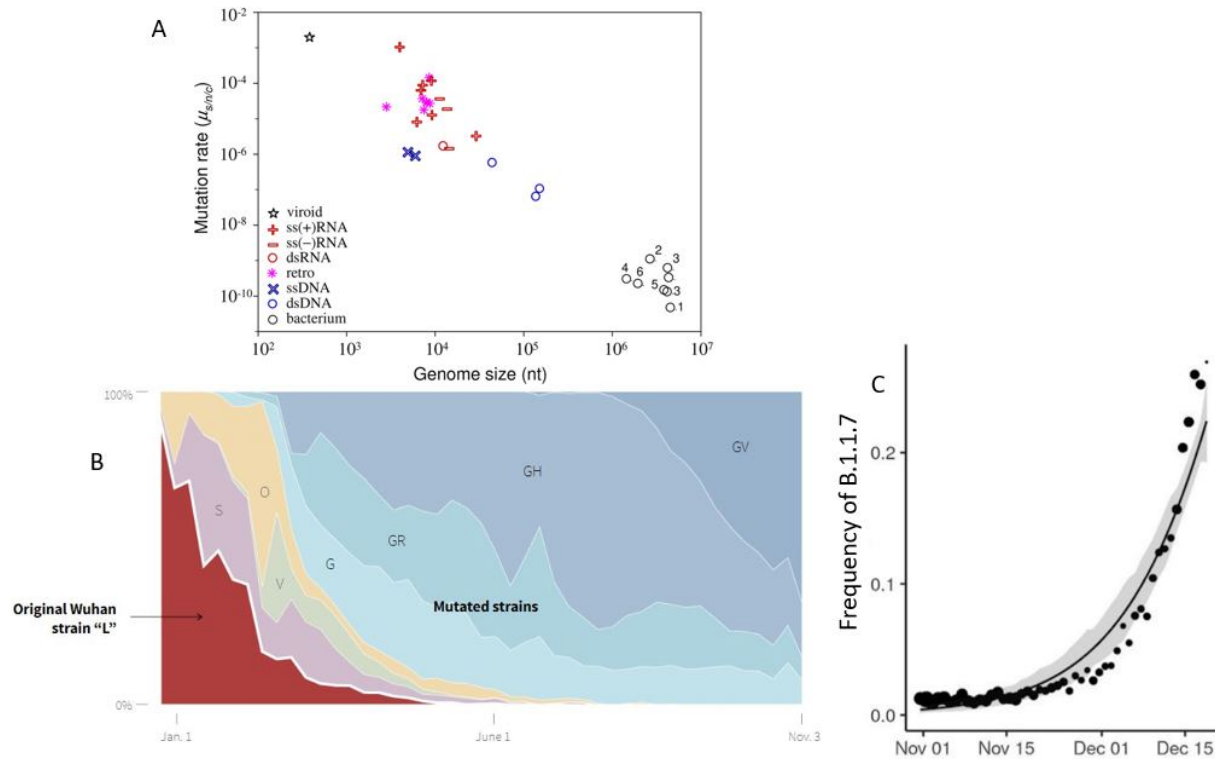
**3.** (1 point) *One speaker discussed how ecologists, archaeologists, and historians can use tree rings to date events based on comparing the relative width of tree rings from trees of known age (i.e., living trees) to those of unknown age (e.g., fallen logs, logs from buildings) to line up dates. Tree rings form due to the differential growth that occurs in a tree between the sunny summer and dark winter. However, both growth and reproduction require energy. The above data come from studies by Knops et al (2007) and Johnson et al (2017) published in PNAS and Forestry. Though the data come from different tree species, assume both patterns are true for all trees for the purposes of this question.* (a) Given the most straight-forward interpretation of the data above, how does heterozygosity relate to fitness in these tree populations? (b) How does inbreeding effect the expected tree ring width? (c) If I compare two populations of trees, one with frequent population size fluctuations and the other with a stable population, which should have larger trunks and why? (d) For part (a) you used the most straightforward read of these data, but what if that isn't the whole story? Describe a specific testable hypothesis that would help you better understand whether or not there are tradeoffs here.



4. (1 point) One speaker discussed local adaptation in plants, and mentioned a common mechanism is by increasing the number of trichomes (small hair-like structures on plants). Trichomes help protect plants from outside forces, especially insect attacks. When an insect attacks a plant, plants release signalling chemicals that volatilize and can be detected by other, nearby plants. Jasmonic acid is a common one. Dr. Ariel Ogran and colleagues (2019) collected wild arugula plants from two populations (a desert and a scrubland population), then planted them in a common garden. They crossed the plants and kept track of the sires and dams for each plant, then planted full siblings separately. Full siblings were then either kept in a normal environment (control) or were exposed to jasmonic acid (treatment), and the density of trichomes on the plants were scored (0 - 5, with 5 being “lots of trichomes”). The results are shown above. In 100 words or fewer, describe effect of different genetics on trichome number in wild arugula.



5. (1 point) One speaker discussed the ecological importance of seed dispersal. The figure above is from a synthesis by Drs. Diana Gamba and Nathan Muchhala (2020) and shows the impact of various aspects of plant biology (including seed dispersal) on the nuclear  $F_{st}$  of plant populations. (a) What combination of mating system, growth form, pollination syndrome, and seed dispersal would you predict makes speciation most likely. (b) State a clear hypothesis that could be tested and that would shed light on why latitude effects  $F_{st}$  in plants. (c & d) Briefly ( $\leq 100$  words) explain why pollination mode would have a larger impact on  $F_{st}$  than seed dispersal for nuclear genes, and what you would *expect* if the study calculated  $F_{st}$  using mitochondrial or chloroplast DNA.



6. (1 point) One of our speakers discussed the diverse forms of RNA and how they impact our lives, with a focus on Sars-Cov-19 near the end. COVID19 is caused by a coronavirus (Sars-Cov-19). In general, viral nucleotide sequences are all highly functional, with little-to-no unnecessary sequences. Figure A above is from Sanjuán et al (2010) showing data on mutation rates for a variety of infectious agents. Coronaviruses have significantly larger genomes than most other viruses, and they also have additional proofreading mechanisms in the proteins that replicate their RNA strands. Figure B above shows the frequency of different Sars-Cov-19 strains globally for much of 2020, while Figure C shows the frequency of the B.1.1.7 variant of the GV strain locally in England for the last month and a half of 2020. (a) Explain the cause of the negative correlation shown above in panel A. (b) Looking at panel B, describe whether the pattern of change in *most* COVID strain frequencies is more similar to what you'd expect under drift or selection, and justify your response. (c) The B.1.1.7 Sars-Cov-19 variant of concern that recently appeared in England has 17 mutations not present in its nearest relative strains, almost ten times the number expected for a new strain. What evolutionary force(s) seem to be shaping its change in frequency over time and how can you tell? (d) How does the number of cases impact the probability of new strains evolving, and why?

7. (1 point) In 100 words or fewer, explain how linkage disequilibrium relates to *genetic drift*.