## Homework 9
due Wednesday March 8, 2017

Do these problems, and email the results (if you can, in a PDF or else in Word .DOC or .DOCX format. PDF is best as I cannot read some .docx files). Like all of the weekly homeworks this one totals 25 points and is due at the end of that day (at midnight).

Important: Because I will not have this finished in time to return your graded homework in class, please also write, near your name **a mailing address, preferably a UW campus PO box number**. I will return the graded homework to that address, and indicate on it also your course point total and grade.

1. (8 points) Suppose that we have a sample of 2 copies of a gene from a population whose effective population size is 1000, a sample of 3 copies from a population whose effective population size is 5000, and a sample of 10 copies from a population whose effective population size is 9,000. These populations are isolated – they do not exchange migrants. Considering the expected coalescent trees of gene copies in these three populations,

   (i) If we look back in time from the samples, which of the three is expected to have the most recent coalescence of a pair of lineages? *Hint: you can just use the expected times back to the first coalescence.*

   (ii) What are the expected times back, in each population, until all the sampled copies have one ancestor?

2. (9 points) In a population of size 100,000, we sample a coalescent tree of 4 copies of a gene. Doing this, suppose that we find this tree (which I generated by sampling from the coalescent distribution for these values):
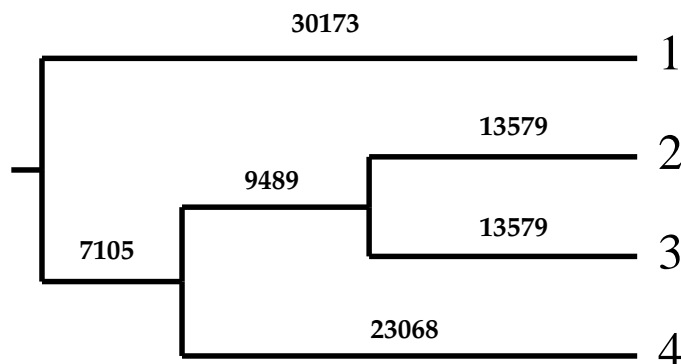
In that tree, the numbers next to each branch are the lengths of each branch in generations.

   (i) How many coalescence events are shown in the tree?

   (ii) For each coalescence, make an estimate of the effective population size N_e by equating the observed length of the coalescence interval with its expectation, and solving for N_e. *(Be careful to figure out, for each one, how many lineages are around when that coalescence is about to occur).*

(iii) It turns out that the best estimate of N_e from all three coalescences is just the average of the estimates in the three coalescence intervals. What value does this have? Compare it to the true effective population size. Is it exactly correct?

(iv) With DNA sequence data, the branch lengths of the tree are not inferred in generations, but in expected mutations per site. If the mutation rate per site is 10^{-8}, what will be the values we see on this tree? What will the overall estimate be?

(v) In that case what quantity are we estimating? *(Hint: it's not simply N_e).*

(vi) In the DNA case, will there be more noise in the estimate than there is if we precisely know the actual coalescent tree and express its branch lengths in expected mutations per site? or less noise? Why?

3. (8 points) For adult body weight in lab, suppose that in an wild population the mean weight is 30g, and the standard deviation of weight from individual to individual is 2g.

(i) If we also collect many parent-offspring pairs and estimate the slope of the regression of offspring on parent, and it is 0.2, what is our estimate of the additive genetic variance of body weight? Why? *(Note: the slope of a regression is the covariance of the two traits, divided by the variance of the trait on the X axis).*

(ii) If we also collect many pairs of full sibs, call the values (x, y), and estimate the slope of the regression of y on x, and it is 0.25, what is our estimate of the dominance variance of adult body weight? Why? *(Careful, you need to think about what the variance of x is expected to be).*

(iii) What is our estimate of the environmental variance? Why?

(iv) What is our estimate of the heritability? Why?