Population Genetics Lab: 26 March 2019

Krimbas and Tsakas (*Evolution* 25:454-460; 1971) studied the evolution of allele frequencies at two different loci encoding soluble esterases in the olive fruit fly, *Dacus oleae*. The product of locus A catalyzes degradation of acetylcholine. The product of locus B catalyzes degradation of pseudocholine. Laboratory experiments had revealed that homozygotes for the null allele at locus A were more susceptible to dimethoate, an organophosphate insecticide, while there was no detectable selection at locus B.

To see whether selection could be detected in wild populations, Krimbas and Tsakas collected flies form an olive orchard in Greece in 1966, 1967, and 1968. A simplified version of the data they collected is summarized in the following table:

	Year		
	1966	1967	1968
Number of A_1 alleles	206	142	230
Sample size of A alleles	474	312	400
Number of B_1 alleles	292	179	280
Sample size of B alleles	469	281	409

Recall that the variance in allele frequency associated with geneetic drift is

$$\mathrm{Var}(p_{t+1}) = \frac{p_t(1-p_t)}{2N_e} \quad . \tag{1}$$

We can use what's known as a Beta distribution to describe allele frequency variation within a population over time. Specifically,

$$\begin{aligned} \mathrm{P}(p_{t+1}|p_t) &\sim \mathrm{Beta}(\nu p_t, \nu(1-p_t)) \\ \nu &= \frac{1-\theta}{\theta} \end{aligned}$$

When the Beta distribution is parameterized in the way illustrated here, it's mean is p_t and its variance is

$$\mathrm{Var}(p_{t+1}) = \theta p_t(1-p_t) \quad . \tag{2}$$

So if we combine (1) and (2), we get

$$\begin{aligned} \frac{p_t(1-p_t)}{2N_e} &= \theta p_t(1-p_t) \\ \frac{1}{2N_e} &= \theta \\ N_e &= \frac{1}{2\theta} \end{aligned}$$

You can use the code in `dacus.R` to run `dacus.jags` and estimate N_e at locus A and locus B. Study the `JAGS` code to make sure you understand what it's doing. Then run it, and think about the following questions:

1. What would your estimate of the effective population size be if you based it on allele frequency changes at locus A?

2. What would your estimate of the effective population size be if you based it on allele frequency changes at locus B?

3. Is there evidence that the estimate of effective population size depends on the locus that you select?

4. What could cause estimates of effective population size to be different at the two loci? Note: You should think about this question even if you conclude in your answer to the previous question that there isn't good evidence that estimates from the two loci differ from one another.

5. Given that the smallest sample included 281 alleles, we know that there are at least 140 individuals in this population. It's reasonable to presume that the sex ratio in this population is close to 50:50. If we assume that generations are non-overlapping and that the variance in individual reproductive success is roughly binomial, are the observed allele frequency changes plausibly accounted for by genetic drift, or might natural selection be involved?

Assume that the changes in allele frequency from year to year are entirely the result of genetic drift, and assume that there is only one generation per year.[1]

[1] There are actually several generations per year, so the estimates for N_e you get below will be substantially smaller than the best estimate of N_e for these data.