

POPULATION GENETICS PROJECT #3

Yoshiko Tobar and Ken-Ichi Kojima (*Genetics* 57:179-188; 1967) studied the evolutionary dynamics of an inversion polymorphism on chromosome 2 of *Drosophila ananassae* in a population cage. Through the standard sort of tricky genetic manipulations that *Drosophila* geneticists do, they derived two lines that were homozygous for each chromosome type. They then started different population cages with differing numbers of flies homozygous for each chromosomal arrangement. Specifically,

Population	Starting configuration	
	AA	BB
1	100	900
2	900	100

After one generation of reproduction in the cage they took a sample of adults and obtained the following “genotype” counts:¹

Population	Genotype		
	AA	AB	BB
1	19	125	156
2	206	87	7

Assume that newly formed zygotes in the generation from which the adult sample was taken are found in Hardy-Weinberg proportions with chromosome frequencies equal to those in the base populations.² Using these data answer the following questions:

1. What are the fitnesses of AA and BB relative to AB in each population, and what are the 95% credible limits on those fitnesses?
2. How likely is it that there is heterozygote advantage in either population?

¹As usual, I’m treating a chromosome inversion type. I’ve also simplified these data a bit. The counts combine results from two replicate populations of each population configuration.

²In other words, the frequency of A in Population 1 is $100/1000 = 0.1$.

3. Are the relative fitnesses of the genotypes the same or different in the two populations?
4. Given your answers to (1) & (2) and assuming that those fitnesses are the only evolutionary force acting, what do you predict about the equilibrium composition of *Drosophila ananassae* populations. Will they be polymorphic? monomorphic for A? monomorphic for B? or will the result depend on initial frequencies? Be sure to explain your answer because (a) if you look up the original paper you will see the result and (b) the pattern of selection happening here is different from anything we've discussed in class.³

Hints

1. To estimate genotype frequencies in the population after one generation of reproduction, you'll need to remember how to estimate proportions in a multinomial. Treat those frequencies as "adult" genotype frequencies. To estimate "zygote" genotype frequencies, you'll need to remember how I constructed the genotype frequencies before selection when I derived the equation for viability selection.
2. Since the JAGS code you'd need to write to estimate fitnesses in each population is identical, you might find it useful to write it once and wrap it all in a `for` loop. In particular, to use the R code I mention a little later, you'll want to use `w.AA[i]` to represent the fitness of genotype *AA* in population *i* ($i = 1, 2$), and `w.BB[i]` to represent the fitness of genotype *BB* in population *i*. You'll probably also want to let `n.a[i,1]` be the count of genotype *AA* in population *i*, `n.a[i,2]` be the count of genotype *AB* in population *i*, and `n.a[i,3]` be the count of genotype *BB* in population *i*. Here's the magic you'll need to chant in R to get the data into that format.⁴

```
n.a <- matrix(nrow=2, ncol=3)
n.a[1,1] <- 19
n.a[1,2] <- 125
n.a[1,3] <- 156
n.a[2,1] <- 206
n.a[2,2] <- 87
n.a[2,3] <- 7
```

³That's not quite true. I mentioned it very briefly. In any case, you have the tools to figure out the results if you think about it for a bit.

⁴Ask Kristen to explain how this works if you don't understand it.

- The genotype counts in the adult sample from population i , i.e., the numbers in `n.a[i]`, represent a multinomial sample from the unknown population genotype frequencies, i.e.,

```
n.a[i,1:3] ~ dmulti(x.a[i,1:3], n[i])
```

where you can calculate `n[i]`, the sample size from population i in R as

```
n <- numeric(2)
for (i in 1:n.pop) {
  n[i] <- sum(n.a[i,])
}
```

- Just to make the notation easier to line up with the notes, I suggest using `x.z[i,1]` to represent the frequency of the AA genotype in newly formed zygotes in population i , `x.z[i,2]` to represent the frequency of AB , and `x.z[i,3]` to represent the frequency of BB .
- Since the genotype frequencies in the adult population may or may not be in Hardy-Weinberg, you'll want to put a prior directly on `x.a[i,1]`, `x.a[i,2]`, and `x.a[i,3]` as follows:

```
for (i in 1:2) {
  for (j in 1:3) {
    alpha[i,j] ~ dexp(1)
    x.a[i,j] <- alpha[i,j]/sum.alpha[i]
  }
  sum.alpha[i] <- sum(alpha[i,])
}
```

- To make it easier to determine the outcome of selection, the code in `process-posterior.R` will extract the posterior distribution of the relative fitnesses from the JAGS output. For the script to work, you'll need to use the notation for relative fitnesses I described above, and you'll want to store the results from the call to JAGS in an object called `fit`. `process-posterior.R` will print a table with the following structure:

	$w_{AA} < 1, w_{BB} > 1$	$w_{AA} < 1, w_{BB} < 1$	$w_{AA} > 1, w_{BB} > 1$	$w_{AA} > 1, w_{BB} < 1$
Population 1				
Population 2				

Numbers in the table will be the number of times the particular fitness configuration was found in the posterior sample for that population. For example, if the element for Population 1 in the column headed $w_{AA} > 1, w_{BB} > 1$ were 500 and the total number of samples in your analysis was 5000, then you'd know that there's roughly a 10 percent chance that both homozygotes have a fitness higher than the heterozygote.