

# BIOB480/BIOE548 notes 10/1/2024

## Introduction

- Reminder: No class Thursday.
- Quiz 2 posted: it's a take-home, instructions on D2L. Due by class on Tuesday (10/8).
- HW5 also due Thursday (10/8). It's a short research and writing assignment—don't be caught off guard.

## Genetic drift recap

Genetic drift is an evolutionary mechanism that causes a change in allele frequencies due to sampling error in the transmission of genetic material. This can be due to either the inherent randomness of “selecting” gametes to form the next generation or variation in survival and fecundity that is unrelated to an organism's phenotype or genotype. (Imagine Wiley E. Coyote standing in the middle of the desert and getting squashed by an anvil through no fault of his own.) Because genetic drift is by definition random (or *stochastic*), its effects on the frequency of a particular allele are unpredictable, unlike the *deterministic* models of mutation, selection, and migration we have been working with up until this point. However, we can make predictions about the likely distribution of allele frequencies in a subsequent generation of size  $N$ . To do so, we make use of the Binomial probability distribution. This probability distribution

$$(p + q)^{2N} = (p + q)(p + q) = p^2 + 2pq + q^2$$

If we have a population that experiences a severe bottleneck—imagine just 2 sexually reproducing diploid organisms survive—we

$$(p + q)^{4N} = (p + q)^4 = (p + q)(p + q)(p + q)(p + q) = p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$$

The generic formula for this probability distribution is as follows:

$$\binom{2N}{K} p^K q^{2N-K} = \frac{2N!}{K!(2N-K)!} p^K q^{2N-K}$$

Here, the first part of the equation—known as the binomial coefficient—is read as “2N choose K”. It expands into a *factorial*, or the product of all positive integers less than or equal to a given positive integer. (Note that both  $0!$  and  $1! = 1$ ) This formula gives the probability of having exactly  $K$  copies of an allele in a diploid population of size  $N$ . In other words, it gives us the probability of a distribution of allele frequencies.

Using our example example of a bottleneck of 2 individuals, here is the probability of ending up with 3 copies of allele  $A_1$  and 1 copy of allele  $A_2$ , assuming the frequencies in the current generation are  $f(A_1) = p = 0.5$  and  $f(A_2) = q = 0.5$ . Note that  $p$  must always indicate the frequency of the focal (not the alternate).

$$\binom{4}{3} 0.5^3 0.5^{4-3} = \frac{4!}{3!(4-3)!} 0.5^3 0.5^1 = \frac{4 * 3 * 2 * 1}{3 * 2 * 1 * (1)} 0.125 * 0.5 = 0.25$$

Genetic drift is stronger in small populations. In addition to randomly changing allele frequencies, it leads to increased homozygosity in individuals, and a population-wide loss of genetic diversity. We turn to that now.

## Genetic drift and diversity loss

To understand how genetic drift impacts genetic diversity, it's helpful to consider an extreme scenario we will return to several times: a population bottleneck, or sudden reduction in population size to a limited number of individuals. The most severe population bottleneck possible for sexually reproducing organisms is 2. Assuming they are diploid, the following table gives us the expected frequencies and heterozygosities of different allelic combinations. ( $H_e$  is calculated as  $2pq$  given  $f(A_1)$  and  $f(A_2)$  from the Allele combination column.)

Allele combination	Frequency ( $f$ )	Heterozygosity ( $H_e$ )	$f \times H_e$
4 $A_1$ s	$p^4$	0	0
3 $A_1$ s; 1 $A_2$	$4p^3q$	0.375	$1.5p^3q$
2 $A_1$ s; 2 $A_2$	$6p^2q^2$	0.5	$3p^2q^2$
1 $A_1$ s; 3 $A_2$	$4pq^3$	0.375	$1.5pq^3$
4 $A_2$ s	$q^4$	0	0
Total	1.0		$1.5pq^*$

\*Derivation: Total =  $1.5p^3q + 3p^2q^2 + 1.5pq^3 = 1.5pq(p^2 + 2pq + q^2) = 1.5pq(1) = 1.5pq$

Mean heterozygosity in the bottlenecked population ( $H_1$ ) will be a proportion of the heterozygosity in the original population ( $H_0$ ). In this case,  $\frac{H_1}{H_0}$  will be  $\frac{1.5pq}{2pq} = \frac{3}{4}$  or  $1 - \frac{1}{4}$ . Since  $2N = 4$ ,  $\frac{H_1}{H_0} = 1 - \frac{1}{2N}$ . Therefore, a single pair bottleneck will reduce  $H_0$  by 25%; a bottleneck of 10 individuals will only reduce it by 5. We will return to this idea from a different perspective in a later lecture.

## Genetic drift activity

See `12_slides.pdf`; `driftR` can be found at <https://cjbattay.shinyapps.io/driftR/>.

## Luikheart and Cornuet Discussion

See `12_slides.pdf`.