

THE HARDY-WEINBERG LAW

We are interested in knowing whether the process of random mating, or more generally, sexual reproduction, can by itself (*i.e.* in the absence of genetic drift and natural selection), cause changes in gene frequencies within a population. The answer to this question is no. To see why this is so, we will derive the Hardy-Weinberg law, which states that random mating does not lead to changes in gene or genotype frequencies.

Consider the following idealized life cycle of an animal resembling a sea urchin: Adults reproduce by shedding gametes more or less simultaneously into the surrounding water. These gametes then combine with each other randomly (hence random mating) to produce diploid zygotes. These zygotes then grow up to be the new generation of adults--or at least some of them do; others will die from various causes (*e.g.* competition, predation, failure to find an attachment site, *etc.*).

We start by considering variation at one locus with two alleles, A and a. At time t , the population is censused to determine the number of individuals of each genotype present, and from these numbers one can calculate, using the formulas given in class, the genotype and gene frequencies in the adults at time t :

Gene and Genotype Frequencies at time t

<u>Genotype</u>	<u>No. Adults</u>	<u>Genotype Freq.</u>	<u>Gene Freq.</u>
AA	N_{AA}	$p_{AA}^t \propto \frac{N_{AA}}{N}$	$p_A^t \propto p_{AA}^t + \frac{1}{2} p_{Aa}^t$
Aa	N_{Aa}	$p_{Aa}^t \propto \frac{N_{Aa}}{N}$	
aa	N_{aa}	$p_{aa}^t \propto \frac{N_{aa}}{N}$	$p_a^t \propto p_{aa}^t + \frac{1}{2} p_{Aa}^t$
<hr style="width: 50%; margin: 0 auto;"/> Total	<hr style="width: 50%; margin: 0 auto;"/> N		

Note that the superscript t indicates the time at which gene frequencies were determined; it **does not** indicate the values of p should be raised to the t^2 power.

Next we want to calculate the number of gametes of each type produced. This is simply the number of individuals of a homozygote genotype times the number of gametes produced per individual (m), or, in the case of heterozygotes, one half of this number:

<u>Adult Genotype</u>	No. Gametes Produced	
	<u>A</u>	<u>a</u>
AA	mN_{AA}	-----
Aa	$\frac{1}{2}mN_{Aa}$	$\frac{1}{2}mN_{Aa}$
aa	-----	mN_{aa}
<hr/>		
Total	$mN(p_{AA}^g + \frac{1}{2}p_{Aa}^g)$	$mN(p_{aa}^g + \frac{1}{2}p_{Aa}^g)$

(remember $N_{AA} \propto p_{AA}N$, etc.)

The total number of gametes is

$$mN(p_{AA}^g + \frac{1}{2}p_{Aa}^g + \frac{1}{2}p_{Aa}^g + p_{aa}^g) \propto mN$$

since the expression in parentheses equals 1. This is as it should be: N individuals are producing m gametes apiece, yielding mN gametes.

From these numbers one may deduce the gene frequencies:

$$p_A^g \propto \frac{mN(p_{AA}^g + \frac{1}{2}p_{Aa}^g)}{mN} \propto p_{AA}^g + \frac{1}{2}p_{Aa}^g, \text{ which from the first table}$$

means that

$$p_A^g \propto p_A^g \text{ (the superscript } g \text{ refers to the gamete stage of the life cycle).}$$

Similarly, one can calculate that $p_a^g \propto p_a^g$.

After gametes have been shed, they combine almost immediately to form zygotes. One can think of the process of zygote formation as being equivalent to choosing first one gamete randomly from the gamete pool, then choosing a second randomly, and finally putting them together to form a zygote. This means that the probability of forming a zygote of a particular genotype, say AA, is simply the probability that the first gamete chosen is A multiplied by the probability that the second gamete chosen is A. These probabilities are simply equal to the proportion of gametes in the gamete pool that are A -- the gametic gene frequency, p_A^g , which

we have already seen is equal to $p_A^>$. Hence, the probability that a zygote will have genotype AA is simply

$$\text{Prob (AA)} = (p_A^>)^2$$

One can in similar fashion deduce the probabilities for the other genotypes, obtaining:

<u>Genotype</u>	<u>Prob. of zygote formation</u> (= <u>Genotype frequencies</u>)		<u>Number of Zygotes</u>
AA	$(p_A^>)^2$	(= p_{AA}^D)	$\frac{1}{2}mN (p_A^>)^2$
Aa	$2p_A^>p_a^>$	(= p_{Aa}^D)	$\frac{1}{2}mN 2 (p_A^>)(p_a^>)$
aa	$(p_a^>)^2$	(= p_{aa}^D)	$\frac{1}{2}mN (p_a^>)^2$
<hr/>	<hr/>	<hr/>	<hr/>
Total	1	(= 1)	$\frac{1}{2}mN$

Note that the probability of obtaining an AA zygote is equal to the zygote genotypic frequency of genotype AA (i.e. $p_{AA}^D = \text{Prob (AA)} = (p_A^>)^2$). A similar relationship holds for the other genotypes.

Note also that the number of individuals having each genotype at the zygote stage is also listed in the above table. This value is obtained by multiplying the value for genotype frequency times the total number of zygotes, which is $\frac{1}{2}mN$, one half the number of gametes.

The final stage of the life cycle is growing up to be an adult. During this process of growth and maturity, there will be heavy mortality. We can specify the proportion of individuals surviving to become adults by l . We assume here that l is the same for all genotypes. Then the number of adults in each genotype surviving to become adults is simply l times the number of zygotes of that genotype:

<u>Genotype</u>	<u>Number of Zygotes</u>	<u>Genotype Freq.</u>	<u>Gene Freq.</u>
AA	$\frac{1}{2}mNl (p_A^>)^2$	$p_{AA}^{>\epsilon} = (p_A^>)^2$	$p_A^{>\epsilon} = p_A^>$
Aa	$\frac{1}{2}mNl 2 (p_A^>)(p_a^>)$	$p_{Aa}^{>\epsilon} = 2p_A^>p_a^>$	
aa	$\frac{1}{2}mNl (p_a^>)^2$	$p_{aa}^{>\epsilon} = (p_a^>)^2$	$p_a^{>\epsilon} = p_a^>$
<hr/>	<hr/>		
Total	$\frac{1}{2}mNl$		

From these numbers, then, one can calculate the genotype frequency at $t \in 1$ for each genotype. For example, for genotype AA, one gets

$$p_{AA}^{>\epsilon''} \propto \frac{N_{AA}^{>\epsilon''}}{N^{>\epsilon''}} \propto \frac{\frac{1}{2}mNl (p_A^>)^{\#}}{\frac{1}{2}mNl} \propto (p_A^>)^{\#} .$$

The values for the other genotypes are listed in the table above.

From the genotype frequencies, one can then calculate the gene frequencies at time $t \in 1$:

$$\begin{aligned} p_A^{>\epsilon''} &\propto p_{AA}^{>\epsilon''} \in \frac{1}{2}p_{Aa}^{>\epsilon''} \propto (p_A^>)^{\#} \in \frac{1}{2}(2)p_A^>p_a^> \\ &\propto p_A^> (p_A^> \in p_a^>) \\ &\propto p_A^> , \text{ since the expression in parentheses is equal to 1.} \end{aligned}$$

In other words, after one generation of random mating, with no selection and no random drift, there is no change in gene frequencies. This result holds for each successive generation as long as the assumptions made initially remain valid.

The situation with genotype frequencies is somewhat more complicated. From the last table, you can see that

$$p_{AA}^{>\epsilon''} \propto (p_A^>)^{\#} .$$

In general, it will *not* be true that, before the first

generation in which our assumptions hold,

$$p_{AA}^> \propto (p_A^>)^{\#} .$$

Consequently, during the first generation of random mating, the genotype frequencies may change. In all subsequent generations, however, there will be no change in genotype frequencies; these will stabilize at the values given in the last table.

We can now state the Hardy-Weinberg law formally:

Under conditions of random mating and no genetic drift or natural selection, no changes in gene frequency will occur; moreover, no changes in genotype frequency will occur after the first generation in which those conditions hold. Finally, under these conditions, genotype frequencies can be calculated from gene frequencies using the "binomial law":

$$(p_A \in p_a)^{\#} \propto p_A^{\#} \in 2p_A p_a \in p_a^{\#} .$$

You can see in the expression above that the genotype frequencies of, respectively, the AA homozygote, the heterozygote, and the aa homozygote, are given on the right-hand side of the equation. This equation can be generalized to a situation in which there are more than two alleles at a locus.

Finally, note that once a population has reached Hardy-Weinberg equilibrium, alleles are associated with each other randomly. To see this, first consider a randomly chosen A allele. Let population size be N. Then there are $2N p_A$ A alleles that are associated with another A allele, while there are $2N p_A p_a$ A alleles that are associated with an a allele. Thus, the proportion of A alleles that are associated with other A alleles is

$$\text{proportion} = \frac{2N p_A}{2N p_A + 2N p_A p_a} = \frac{p_A}{p_A + p_a} = p_A .$$

Next, consider a randomly chosen a allele. There are $2N p_a$ of these alleles that are associated with an A allele, while there are $2N p_a^2$ that are associated with another a allele. Thus, the proportion of a alleles that are associated with A alleles is

$$\text{proportion} = \frac{2N p_a}{2N p_a + 2N p_a p_A} = \frac{p_a}{p_a + p_A} = p_a ,$$

i.e. the same proportion as for A alleles. In other words, A and a alleles each have the same probability of being associated with an A allele (and they have the same probability of being associated with an a allele). This means the alleles are associated randomly in diploid genotypes. It can thus be said that **the process of random mating randomizes the associations among alleles at a single locus.**