

## CHANGE IN GENE FREQUENCY DUE TO NATURAL SELECTION

This handout derives a general equation for changes in gene frequency due to natural selection. Consider a single locus with two alleles, A and a, and an organism with a simple life cycle, such as the urchin discussed in the Hardy-Weinberg handout. Suppose we start by enumerating all the individuals of each genotype at this locus just after zygotes are formed. Let  $p^t$  be the gene frequency of allele A at this time, let  $q^t = 1 - p^t$  be the frequency of allele a, and let  $N^t$  be the number of zygotes in the population at this time. Note that the superscript  $t$  indicates a particular time and does not indicate a power. Assuming random mating, we then have:

<u>Genotype</u>	<u>Frequency</u>	<u>No. zygotes</u>
AA	$(p^t)^{\#}$	$N^t(p^t)^{\#}$
Aa	$2p^tq^t$	$2N^tp^tq^t$
aa	$(q^t)^{\#}$	$N^t(q^t)^{\#}$

Next, the zygotes grow and develop and turn into adults. However, because there are a lot of dangers along the way, most will not survive to become adults. Let  $l_{AA}$ ,  $l_{Aa}$ , and  $l_{aa}$  be the proportions of zygotes surviving to the adult stage for genotypes AA, Aa, and aa respectively. Note that when we derived the Hardy-Weinberg law, we assumed these quantities were all equal. Now we allow them to differ to reflect differences among genotypes in mean probability of survival (*i.e.* natural selection due to differential survival). The number of zygotes of each genotype surviving is then given by the second column in the table below:

<u>Genotype</u>	<u>No. adults</u>	<u>No. gametes produced</u>
AA	$l_{AA} N^t(p^t)^{\#}$	$2m_{AA} l_{AA} N^t(p^t)^{\#}$
Aa	$l_{Aa} 2N^tp^tq^t$	$2m_{Aa} l_{Aa} 2N^tp^tq^t$
aa	$l_{aa} N^t(q^t)^{\#}$	$2m_{aa} l_{aa} N^t(q^t)^{\#}$

The next phase of the life cycle is the production of gametes. Allowing for differential fertility, we can represent the number of gametes produced by an adult of each genotype by  $2m_{AA}$ ,  $2m_{Aa}$ , and  $2m_{aa}$ . Multiplying this by the number of adults of the appropriate genotype gives the total number of gametes produced by each genotype, column 3 in the table above.

At this stage, we can calculate the gene frequencies in the gametes. The gene frequency of allele A is simply the number of gametes that carry allele A divided by the total number of gametes:

No. A gametes  $\propto$  (No. gametes produced by AA)  $\propto \frac{1}{2}$  (No. produced by Aa)

$$\propto 2m_{AA} l_{AA} N^2(p^2)^{\#} \propto \frac{1}{2} 2m_{Aa} l_{Aa} 2N^2 p^2 q^2$$

Total No. gametes  $\propto 2m_{AA} l_{AA} N^2(p^2)^{\#} \propto 2m_{Aa} l_{Aa} 2N^2 p^2 q^2 \propto 2m_{aa} l_{aa} N^2(q^2)^{\#}$

Therefore, the gene frequency of allele A is

$$p_A \propto \frac{2m_{AA} l_{AA} N^2(p^2)^{\#} \propto \frac{1}{2} 2m_{Aa} l_{Aa} 2N^2 p^2 q^2}{2m_{AA} l_{AA} N^2(p^2)^{\#} \propto 2m_{Aa} l_{Aa} 2N^2 p^2 q^2 \propto 2m_{aa} l_{aa} N^2(q^2)^{\#}}$$

$$\propto \frac{p^2 \cdot m_{AA} l_{AA} (p^2)^{\#} \propto m_{Aa} l_{Aa} q^2}{m_{AA} l_{AA} (p^2)^{\#} \propto 2m_{Aa} l_{Aa} p^2 q^2 \propto m_{aa} l_{aa} (q^2)^{\#}}$$

Next, we define the *fitnesses* of the three genotypes as

$$W_{AA} \propto m_{AA} l_{AA}$$

$$W_{Aa} \propto m_{Aa} l_{Aa}$$

$$W_{aa} \propto m_{aa} l_{aa} .$$

The fitnesses represent the expected number of gametes a genotype will produce, taking mortality into account. Substituting these fitnesses into the above equation yields a general equation for gene frequency change due to natural selection:

$$p^{2\#} \propto \frac{p^2 \cdot p^2 W_{AA} \propto q^2 W_{Aa}}{(p^2)^{\#} W_{AA} \propto 2p^2 q^2 W_{Aa} \propto (q^2)^{\#} W_{aa}}$$

The denominator of this expression is just the mean fitness in the population (the weighted sum of the fitness of each genotype, where the weights are the genotype frequencies). The expression in brackets in the numerator is known as the *marginal fitness* of the A allele. It is the mean fitness of individuals carrying at least one A allele. The new gene frequency is thus the old gene frequency times the ratio of marginal fitness to mean fitness.