

Topic 16. The Process of Evolution and Population Genetics

I. Darwin's Theory of Evolution

Charles Darwin and **Alfred Russel Wallace** jointly proposed the theory of evolution in 1858. Each had independently reached the conclusion that species change over time, and that this process generates new species. Darwin had sat on his ideas for over a decade only to be prompted to publish through his encounter with Wallace's unpublished work. They were not the first to propose that organisms evolved into new forms, but they were the first to postulate the idea with a mechanism to explain these changes. Darwin, in particular, was influenced by the works of **Thomas Malthus**, and the geologists, **James Hutton** and **Sir Charles Lyell**. Malthus proposed that, because population growth increases exponentially, and food production only linearly, that societies would always teeter on the brink of famine. From this idea, Darwin concluded that, for any given species, more individuals would be born than could possibly be supported by the finite resources of their environment. Exceptions would occur only when a population is exploiting a completely open environment, an extremely rare circumstance. From Hutton and Lyell, Darwin concluded that the age of the Earth must be measured at least in millions of years. From these ideas, Darwin surmised that survival of individuals in each generation would not be a random event, and that, given the age of the Earth, a huge number of generations must have transpired in the past, it was a logical step to conclude that the environment would affect heritable characters of organisms in subsequent generations through the removal of those members of the preceding generation least fit to survive.

(for more information on Malthus visit <http://www.ucmp.berkeley.edu/history/malthus.html>)

(for more information on Hutton visit <http://www1.umn.edu/ships/religion/hutton.htm>)

(for more information on Lyell visit <http://www.victorianweb.org/science/lyell.html>)

To support this idea, Darwin used examples of species modified by human selection. He took up the hobby of breeding pigeons. The breeds of pigeon at the time were dramatically different from one another in both appearance and behavior and yet all could interbreed. The variation between these breeds superficially was greater than that found between many species. Darwin concluded that all the breeds had arisen from a wild progenitor but had diverged due to selective breeding. Similar species, he surmised, were derived from a common species in the past in much the same way. The difference being, that, in nature, selection was determined by the environment, and not by the decisions of a breeder. Given the age of the Earth, the next logical step was to surmise that all species had been derived from the first living organism to arise on Earth.

Darwin's ideas were incomplete due to the immature state of the science of genetics. It was not understood how the observable characters of organisms were carried from one generation to the next, nor how this information was expressed in the individual. Also left unanswered was the question of how variation was introduced into populations. Natural selection could weed out certain characters, but there was no evidence that it could actually generate anything new. Without an understanding of how variation arose, it was impossible to explain the range of differences seen between living species.

The Modern Synthesis of the Twentieth Century

It was only in the early twentieth century that the discoveries of Gregor Mendel became commonly known throughout the scientific community. It was then clear how heritable characters were passed from one generation to the next. Further, the geneticists of the time became aware of the phenomenon of **mutation**. Organisms that are pure breeding can spontaneously express something new due to a change in the underlying genetic information. At the time, biologists incorporated the discoveries from genetics into Darwin's ideas to generate a synthesis of the two. By this synthesis, evolution is best considered as a change in the genetic makeup of a population, which in turn, is seen as changes in the phenotypes of a population over time. The source of the variation necessary for evolution is mutation. These ideas can be more fully explored through the subdiscipline of genetics termed population genetics.

II. Population Genetics

When considering genetics at the population level, we must determine the numbers of copies of each allele found at a locus under consideration in a population. This constitutes a subset of the gene pool. The gene pool includes all the copies of all the alleles at every locus found in a population (the scope of which is a bit too grand to treat all at once). Remember that alleles are carried in an organism two at a time. If we tally all the copies of a specific allele for a locus, and divide that number by all the copies for every allele, this gives us the **allelic frequency**.

Example: For a species of a hypothetical plant, the trait flower color is determined by a pair of alleles: **R** = red flowers and **r** = white flowers. If we have a population of 100 plants with the genotypes listed below, what is the allelic frequency for the "R" allele? : for the "r" allele?

Genotype	RR	Rr	rr
Number of individuals	36	48	16

$$\text{The frequency of the "r" allele} = [48 + (16 \times 2)] \div 200 = .4$$

$$\text{The frequency of the "R" allele} = [48 + (36 \times 2)] \div 200 = .6$$

Note that the sum of all the frequencies for a given locus must equal 1

$$.4 + .6 = 1$$

We can use these frequencies to determine the genotypes and their frequency in the next generation **if we assume that mating is random**. Any given pollen grain collected from the population as a whole has a 40% chance of carrying the allele for white flowers. Any given ovule also has a 40% chance of carrying the allele for white flowers. The odds of these two independent events coming together is their product:

$$.4 \times .4 = .16$$

Thus, the chance that any random individual of the next generation will have the genotype “rr” is .16, and the frequency of that genotype in the next generation should be .16. By the same logic the frequency of the genotype “RR” should be .36. We can consider the frequency for all possible genotypes in the next generation through the use of a modified punnet square. In this example we consider not only the possible genotypes of pollen grains and eggs, but also their allelic frequency. In each cell is entered the combination of the two events: the resulting genotype and its frequency is simply the product of the two allelic frequencies of the combining alleles.

.4 r	.16 rr	.24 rR
.6 R	.24 Rr	.36 RR
	.4 r	.6 R

When we tally the result for a population of 100 individuals, we see that we end up with a new generation with the same allelic frequency as the original:

36 RR: 48 Rr: 16 rr.

This is consistent with the rule in population genetics called the **Hardy-Weinberg Law**.

Hardy-Weinberg Law

The Hardy Weinberg Law states that allelic frequencies in a population will remain the same from one generation to the next unless factors other than sexual recombination come into play. In any generation, the frequency of genotypes of

individual members can be determined as illustrated above with a punnet square, or more simply, by expanding the expression

$$(p + q)^2$$

Where “q” = the frequency of one allele and “p” the frequency of the second allele (assuming there are only two alleles found at that locus).

Note that

$$p + q = 1$$

and that

$$(q + p)^2 = q^2 + 2pq + p^2 = 1$$

Problem 1: On Pitcairn Island in the South Pacific, all four blood types “A”, “B”, “AB”, and “O” are found in the population. Emigration and immigration is virtually nil.

The frequency of the alleles for blood type are

$$\begin{aligned} \text{“A”} &= .3 \\ \text{“B”} &= .2 \\ \text{“O”} &= .5 \end{aligned}$$

What proportion of the population has blood type, “A”?

What proportion of the population has blood type, “B”?

What proportion of the population has blood type, “AB”?

What proportion of the population has blood type, “O”?

Factors that Change Allelic Frequencies in a Population:

There are five factors listed in your book that can change this equilibrium between generations, resulting in evolutionary change.

1. **Mutations.**
2. **Emigration or immigration.**
3. **Selective mating.**
4. **Small population size, where chance results differences in success .**
5. **Natural selection.**

Problem 2: Pitcarn island was settled by the mutineers of the HMS Bounty together with their Polynesian wives. The allelic frequencies for blood type listed for Pitcarn Island is neither like that of the English population as a whole, nor like that for Polynesians native to Tahiti, neither is it simply an averaging of the two. Which factors listed above could explain the current allelic frequencies found there?

Mutation

A mutation at a locus generates new alleles from some original gene. For example the new allele may encode for a defective protein such as an enzyme in some biosynthetic pathway. The “R” allele introduced earlier may encode for one enzyme in a pathway leading to anthocyanin (the red pigment in flowers) and the “r” allele was derived by an error in the replication of “R”. The “r” allele results in white flowers because the protein it encodes is defective. Further, it is recessive because the plant only requires one copy of the “R” allele to produce this enzyme. Indeed, alternate alleles for a given locus were derived by mutation. In the absence

of natural selection, over hundreds of generations, even a slow rate of mutation will alter the allelic frequencies at a locus.

Will all base changes in a sequence of DNA encoding for a protein result in a change in the function in the protein encoded?

Emigration and Immigration

With animals this is a fairly obvious, but is a bit more subtle for plants rooted to the ground. Based on your knowledge how pine pollen is dispersed, can a population of white pine in Tennessee be completely isolated from one in Arkansas?

How about a population of white pine in Arkansas from the one in Tennessee?

Selective Mating

In animal populations, females can drive the evolution of secondary sexual characteristics by favoring certain characters in the males. This can result in the evolution of features such as the plumage of peacocks that has no other selective value.

Can you think of situations where pollination in a plant population will fail to be random?

Natural Selection

Natural selection can act on a population by favoring the survival and reproduction of certain characters over others. Natural selection only acts on phenotypes and not genotypes. Hence, for our flower color example, the genotypes “RR” and “Rr” genotypes are equally fit as they both have red flowers. In the absence of complete dominance, each genotype may result in a different phenotype

each with a different degree of fitness. In theory, we can quantify the relative viabilities for different genotypes (based on their phenotypes), but actually determining these values is next to impossible. Still this is a useful concept for illustrating how natural selection can drive evolution.

The viability of a genotype is tied to its phenotype and is simply the relative probability that an individual of that genotype will survive to reproduce. Consider two alleles “A₁” and “A₂” found in a hypothetical population. Possible genotypes for this locus are

A₁A₁

A₁A₂

A₂A₂

Viability is represented by “w”, and the viability for each genotype can be symbolized as follows:

Viability of A₁A₁ = w₁₁

Viability of A₁A₂ = w₁₂

Viability of A₂A₂ = w₂₂

Note: if we are dealing with complete dominance then w₁₁ = w₁₂.

If the frequency of each genotype is “x”, “y” and “z” before selection such that

x = frequency of A₁A₁ genotype

y = frequency of A₁A₂ genotype

z = frequency of A₂A₂ genotype

Then the relative frequencies after selection will be

frequency of A₁A₁ genotype = (x)(w₁₁)

frequency of A₁A₂ genotype = (y)(w₁₂)

frequency of A₂A₂ genotype = (z)(w₂₂)

If w₁₁ = w₁₂ > w₂₂ what will happen to the allelic frequency of A₂ in the next generation?

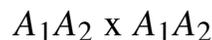
If the relative viabilities remain the same over a series of generations, then the ultimate fate of each allele can be predicted based on the contingencies listed below:

$w_{11} > w_{12} > w_{22}$	A_1 fixes (achieves allelic frequency of 1.0) , A_2 is lost
$w_{11} < w_{12} < w_{22}$	A_1 is lost, A_2 fixes (achieves allelic frequency of 1.0)
$w_{11} < w_{12} > w_{22}$	A_1 and A_2 both protected from loss
$w_{11} > w_{12} < w_{22}$	A_1 or A_2 fixes, depending on initial frequencies

The first two cases are called directional selection. The third is called balancing selection, or over-dominant selection. The fourth is called disruptive selection, or under-dominant selection. A fifth possible case is that there is no selection at all ($w_{11} = w_{12} = w_{22}$). This is called neutrality; we say that allele A_1 and allele A_2 are neutral. Strictly speaking, the four rules above apply only when the population is infinite in size. When, as must be true in nature, the population is finite in size, selection pushes alleles in the directions indicated above, but now this deterministic force can be counteracted by random events.

Evolution Due to a Small Population Size

Even in the absence of selective differences, if the population size is finite, alleles will change in frequency from one generation to the next. This occurs because, through chance, some individuals will happen not to reproduce while others will leave more than their fair share of offspring. It is also the result of chance as to which genotypes are actually produced by individuals that successfully reproduce. This phenomenon is known as genetic drift. Genetic drift is a random process. For example, consider the case of a population of 1 organism ($N = 1$) where the single individual has genotype A_1A_2 . This individual must pollinate itself. This cross can be written



From Mendel's first law we should expect 1/4 of the progeny to have the genotype, A_1A_1 , 1/2 to have A_1A_2 and 1/4 to have A_2A_2 . If only one offspring were to survive, then there is a 25 per cent chance that it will be A_1A_1 in which case the allele A_2 will be lost. There is also a 25 per cent chance that the offspring will be A_2A_2 in which case the allele A_1 will be lost. There is only a 50 per cent chance that both alleles will be preserved in the gene pool of the next generation

The model, which is generally used to approximate genetic drift in evolutionary and population genetic studies, is called the Wright-Fisher model. This model is named after two of the founders of theoretical population genetics: Sewall Wright (a professor at U.W.-Madison), and Sir Ronald A. Fisher. Not coincidentally, R. A. Fisher was also one of the founders of statistics. By this model, the genetic variation within a population of “N” individuals breeding at random will, on average, be halved in $1.4 N$ generations through the loss of some alleles and the fixation of others (fixation is where the frequency of an allele is 1.0).

What to Know for the Quiz

The last four pages will be submitted to your TA for grading. This is inserted here so you may review concepts when these four pages are unavailable to you, and, hopefully, will also serve to focus your attention on important concepts.

1. Understand what the following variables represent in population genetics, as introduced here, and be able to apply them: 'p', 'q', 'A₁', 'A₂', w₁₁, w₁₂, w₂₂ and 'N'.
2. Know what it means for an allele to become 'fixed'.
3. Know what it means for an allele to become 'extinct'.
4. Know how population size effects the odds of an allele becoming fixed or extinct.
5. Know how the gene frequency effects the odds of an allele becoming fixed or extinct.
6. Understand the relative importance in preserving an allele, of its gene frequency, and of the population size.
7. Know the numerical relationship between 'p' and 'q'.
8. Be able to derive gene frequencies given genotype frequencies of a population.
9. Be able to derive genotype frequencies of a population given gene frequencies.
10. Know the Hardy-Weinberg Law, and the conditions that must be met for gene frequencies to remain unchanged over time. Understand specifically how these conditions are violated in real populations.

Computer-based Simulations for Natural Selection and Genetic Drift

Work in pairs to complete the tasks in Part I. At the end of the lab, write both names on your results and turn them in. Be sure to take turns at this: one person clicking the mouse and one person recording the results. Part II requires the results of all pairs to be pooled. Each individual must complete the problems in Part II on her/his own, and turn them in at the end of lab.

Open the simulations directory by dragging the file 'index' onto the Internet Explorer icon on your desktop.

Part I

Task 1: Follow the link to Natural selection on the index page.

Try the following five conditions and report the final allele frequency. Also write in words: e.g., "allele 1 fixed," "allele 1 and 2 both maintained," "allele 2 fixed."

a) $p = 0.1$, $w_{11}=1.2$, $w_{12}=1.0$, $w_{22}=0.8$

Allele frequency: _____

In words: _____

b) $p = 0.1$, $w_{11}=0.8$, $w_{12}=1.1$, $w_{22}=1.2$

Allele frequency: _____

In words: _____

c) $p = 0.1$, $w_{11}=0.8$, $w_{12}=1.1$, $w_{22}= 0.9$

Allele frequency: _____

In words: _____

d) $p = 0.1$, $w_{11}=1.1$, $w_{12}=0.9$, $w_{22}=1.1$

Allele frequency: _____

In words: _____

e) $p= 0.9$, $w_{11}=1.1$, $w_{12}=0.9$, $w_{22}=1.1$

Allele frequency: _____

In words: _____

Task 2: Follow the link to Genetic drift on the index page.

Set “Generations” to 250. For each of the four-parameter sets below, run the simulation fifty times and record the number of times allele A1 fixes. If the simulation ends before either fixation or loss occurs, record a fixation event if the frequency of A1 is greater than one half.

(a) $p_1 = 0.1, N = 10$

Number of times A1 goes to fixation _____ Frequency = _____

(b) $p_1 = 0.1, N = 25$

Number of times A1 goes to fixation _____ Frequency = _____

(c) $p_1 = 0.5, N = 10$

Number of times A1 goes to fixation _____ Frequency = _____

Task 3: Follow the link to Natural selection and genetic drift on the index page.

Note that the default fitness here are $w_{11} = 1.0, w_{12} = 0.9,$ and $w_{22} = 0.8.$ Set “Generations” to 250. For each of the two-parameter sets below, run the simulation fifty times and record the number of times allele A1 fixes. If the simulation ends before either fixation or loss occurs, record a fixation event if the frequency of A1 is greater than one half.

(a) $p_1 = 0.1, N = 25$

Number of times A1 goes to fixation _____ Frequency = _____

(b) $p_1 = 0.1, N = 100$

Number of times A1 goes to fixation _____ Frequency = _____

Task 4: The TA will gather the results from each lab pair. After getting these, calculate the means results for tasks 2-3.

Part II

Provide a brief answer to each of the following questions. Use the results from Part I to support your answers.

Problem 1: How does the probability of fixation of an allele in a finite population depend on the starting allele frequency when there is no selection?

Problem 2: How does the probability of fixation of an allele depend on the population size when there is no selection?

Problem 3: How does the probability of fixation of a selectively favored allele depend on the population size?

Problem 4: What happens in an infinite population when $w_{11} > w_{12} < w_{22}$ with $w_{11} = w_{22}$ and a starting allele frequency of 0.5? What happens in a real (finite) population under these conditions?