Random genetic drift

- When the population size is not very large, frequencies of alleles can change by random events.

- The random change in allele frequencies occurs in nature, because only a limited number of adult individuals can be produced, even though their parents can produce a much larger number of gametes.

- In this toy model, we assume that a locus A in a haploid has two alleles $A_1$ and $A_2$, the population have a fixed 10 individuals in each generation, and they generate a vast number of gametes with the same allele frequencies as in the parents generation.

<table>
<thead>
<tr>
<th>Generation</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>0.7</td>
</tr>
</tbody>
</table>

- $A_1$, $A_2$
Random genetic drift

- The process of change in allele frequency solely due to chance effects is called **random genetic drift**.

- We can study random genetic drift quantitatively by making a few assumptions to simplify the problem:
  1. Ignore nature selection, so all individuals have the same fitness, i.e., the mutant alleles are neutral. Therefore, they are chosen with a probability equal to their proportion in the population.
  2. There is no overlap between generations, i.e., individuals are reproduced and die simultaneously.
  3. The population size is fixed, so, if there are \( N \) individuals in the population, we have \( N \) copies of alleles at a locus for a haploid, and \( 2N \) copies of alleles at a locus for a diploid, .

- Let us consider a locus with two alleles \( A_1 \) and \( A_2 \), with frequencies \( p \) and \( q=1-p \), respectively. The number of \( A_1 \) alleles in the next generation, denoted by \( X \), is a random number, and it follows a **binomial distribution**.
Random genetic drift

- For a haploid, the probability of having \( X = x \) \( A_1 \) alleles in the next generation will be:

\[
P(X = x) = \binom{N}{x} p^x (1 - p)^{N-x}, \quad p = \frac{m}{N}.
\]

where \( m \) is the number of allele \( A_1 \) in the parent generation.

- We can also study random genetic drift using **computer simulations** by starting with a fixed number \( N \) of alleles, and randomly choosing \( N \) alleles for the next generation with replacement.

- The simulation will be repeated by a large number of times (e.g., 1,000 replicates), then statistics can be computed for each generation from these simulations.
Computer simulation of random genetic drift

- Probability distributions of allele frequencies in a diploid population of $N=10$ with starting frequencies $p = 0.3$ and $q = 0.7$. Results are the average of 1,000 simulations, and white bars are theoretic results.
Computer simulation of random genetic drift

- Representative simulations of the change in frequency of allele $A_1$ in diploid populations of $N=25$ and $250$, with starting frequencies $p = q = 0.5$. 

![Graph showing allele frequency over generations for populations of $N=25$ and $N=250$.]
Computer simulation of random genetic drift

- Representative simulations of the fate of newly arising allele in a haploid population with $N=200$ and $p = 1/200$. 

![Graph showing fixation and extinction of a neutral mutation over time.](image_url)
Random genetic drift of an advantageous mutant allele

- If the $A_1$ mutant allele has a selection advantage $s$ (i.e., its fitness is $1+s$, $s>0$), then the probability that the allele will appear in the next generation is proportional to its fitness ($m$ is the initial number of mutants):

$$p = \frac{m(1+s)}{N \bar{w}},$$

where $\bar{w}$ is the mean fitness of the mixed population,

$$\bar{w} = \frac{m(1+s) + (N-m)}{N}.$$

- The probability of having $i$ $A_1$ alleles in the next generation will still follow a binomial distribution with $p$ defined above.

![Graphs showing the probability of having $A_1$ alleles over time for different values of $s$.]
Gene substitution

Gene substitution is the process whereby a mutant allele completely substitutes the predominant or the original wild type allele in a population.

In this process, a mutant allele often arises as a single copy, and then it becomes fixed after a certain number of generations.

There are three related issues associated with the analysis of gene substitution:

1. Fixation probability \( (p_{fix}) \): the probability that a new mutant allele will become fixed.

2. Fixation time \( (t) \): the time it takes for a new allele to become fixed. We usually consider the mean fixation time \( (\bar{t}) \) of a mutant.

3. Rate of gene substitution \( (K) \): the number of fixations per unit time.
Gene genealogical tree and gene coalescence

- A tree that shows the parent-child relationships of genes during evolution/history is called a genealogical tree of the gene.

- It has been theoretically shown and experimentally demonstrated that if we go back enough number of generations, all existing descendant genes can be traced back to a single most recent common ancestor (MRCA) gene.

- This phenomenon is called gene coalescence.

That is, gene mutation always arises from a single mutation.

- Gene coalescence is inevitable simply because not all individuals in any generation have an offspring.
Evidence for gene coalescence

Analysis of mitochondrial genomes of human individuals from different parts of the world, suggests that there was a common ancestral mitochondrial genome originated in Africa about 170,000 years ago. The similar conclusion has been drawn for the Y-chromosome.

The polymorphic sites from the sampled mitochondrial genomes are arranged in a decreasing order of the number of alleles in the sites, not according to their positions on the genome.
A model of the coalescence process

Assume that 1) the population has a fixed size $N$ in each generation during the course of evolution, 2) all individuals are equally fit (only neutral mutations are allowed), and 3) the population mate randomly.

Then the probability of two individuals in the present generation having the same mother in the last generation is:

$$ p = \frac{1}{N^2} N = \frac{1}{N}. $$

The probability that they do not have the same mother is $1 - \frac{1}{N}$.

The probability that two individuals have the same mother $T$ generations ago is,

$$ P(T) = (1 - \frac{1}{N})^{T-1} \frac{1}{N} $$

Geometric distribution

$$ \approx \frac{1}{N} e^{-T/N} = \lambda e^{-\lambda T}, \quad \lambda = \frac{1}{N}. $$

Exponential distribution

$$(e^{-\lambda} \approx 1 - \lambda)$$

Therefore, the average time needed to trace back, so that two individuals had the same ancestor is $N$ generations.

It can be shown that the average time till coalescence of all $N$ individuals is $2N$ generations.
**Gene fixation probability**

- **Fixation probability of neutral mutants:** If a haploid population has a size of \( N \) individuals, then the probability for a neutral mutant at a locus to become fixed is \( p_{\text{fix}} = 1/N \).

- This is because, although any copy of neutral mutant allele in the ancestor generation can be possibly fixed in the current generation, however, according to the gene coalescence theory, only one of such parents can transfer its allele to the current generation, therefore the probability for one of this mutant to become fixed is \( 1/N \).

- **Fixation probability of advantageous mutants:** If a mutation has a fitness \( w = 1 + s, \ s > 0 \) then its fixation probability in a haploid population of size \( N \) is

  \[
  p_{\text{fix}} = \frac{1 - e^{-2s}}{1 - e^{-2Ns}}.
  \]

- **Fixation probability of deleterious mutants:** If a mutation has a fitness \( w = 1 - s, \ s > 0 \) then its fixation probability in a haploid population of size \( N \) is

  \[
  p_{\text{fix}} = \frac{e^{2s} - 1}{e^{2Ns} - 1}.
  \]
Gene fixation probability

- Fixation probability as a function of selection coefficient in a haploid population of size $N=200$.

$$p_{fix} = \frac{1 - e^{-2s}}{1 - e^{-2Ns}}.$$
Gene fixation probability

- If \( s \to 0 \), i.e., the mutation is near neutral, then from \( p_{\text{fix}} = \frac{1 - e^{-2s}}{1 - e^{-2Ns}} \), we have,

\[
p_{\text{fix}} = \frac{1 - e^{-2s}}{1 - e^{-2Ns}} \approx \frac{1 - (1 - 2s)}{1 - (1 - 2Ns)} = \frac{2s}{2Ns} = \frac{1}{N}. \quad \text{(Because } e^{-x} \approx 1-x)\]

This conclusion is the same as we derived from the gene coalescence theory.

- If \( s > 0 \), but is small as in most cases happening in nature, i.e., slightly positive selection, and \( N \) is large, then,

\[
p_{\text{fix}} = \frac{1 - e^{-2s}}{1 - e^{-2Ns}} \approx \frac{1 - (1 - 2s)}{1 - 0} = 2s.\]

- Since when \( s \) is small, \( 2s \) is also small, therefore, advantageous mutations can still have a good chance not being fixed.

- Other the other hand, when \( s \) is small, deleterious mutation can still have a chance to become fixed, albeit such a chance is even smaller. With the increase in deleterious level of a mutation, the chance of its fixation drops very rapidly.
Gene fixation time

- The mean fixation time for all mutants that will eventually become fixed in the population is called **conditional fixation time**.

- **Gene fixation time for a neutral mutant:** In the case of a new neutral mutation in a diploid population, the conditional fixation time is approximated by \( \bar{t} = 2N \) generations.

- **Gene fixation time for a advantageous mutant:** For a mutation with a selective advantage of \( s \), it is approximated by \( \bar{t} = \frac{2/s}{\ln(2N)} \).

- **Gene fixation time for a deleterious mutant:** Interestingly, the conditional fixation time for a mutation with a selective disadvantage of \(-s\), is also approximated by \( \bar{t} = \frac{2/s}{\ln(2N)} \).

This is understandable given the low fixation probability of deleterious mutations.

- Clearly, \( 2N > \frac{2/s}{\ln(2N)} \), i.e., it takes a longer time for a neutral mutant to be fixed.
Gene fixation time

In other words, both advantageous and deleterious mutations are either rapidly lost or fixed, but change in the frequency of neutral mutations is rather slow, thus a large amount of transient polymorphism is generated.
Rate of gene substitution

- **Rate of gene substitution** (\(K\)) measures how often an allele is replaced by a new mutant, and is defined as the **number of fixations occurred at the locus per unit time** (measured in million years).

- **Rate of gene substitution for neutral mutations:**

  For a haploid species of population size \(N\), if a locus mutates at a rate of \(u\) (number of mutations per time unit), then there are \(uN\) mutations in the population at this locus. Since the fixation probability for neutral mutations are \(1/N\), the rate of substitution is

  \[ K = uNp_{fix} = uN \frac{1}{N} = u. \]

- Therefore, for neutral mutations, the **rate of substitution is equal to the rate of mutation**, a remarkably simple result.

- \(K\) is also independent of the populations size: in a large population, the number of mutations arising is a large, but their fixation probability is low; in a small population, the number of mutations arising is a small, but their fixation probability is high.
Rate of gene substitution for advantageous mutations:

The fixation probability of advantageous mutations \((s > 0)\) is

\[
K = uN p_{fix} = uN \frac{1 - e^{-2s}}{1 - e^{-2Ns}}
\]

When \(s\) is small, \(p_{fix} = 2s\), the rate of substitution is,

\[
K = uN p_{fix} = uN 2s = 2usN.
\]

Therefore for advantageous mutations, the rate of substitution depends on the population size, selection coefficient and mutation rate.

Rate of gene substitution for deleterious mutations will be,

\[
K = uN p_{fix} = uN \frac{e^{2s} - 1}{e^{2Ns} - 1}
\]

It also depends on the population size, selection coefficient and mutation rate, but it decreases rapidly with the increase of the deleterious level, \(s\).
Neutral theory of molecular evolution

- **Darwinism:** Darwin proposed his theory of evolution by natural selection without the knowledge of the genetic basis of evolution and the source of variation in populations.

- **Synthetic theory of evolution or neo-Darwinism:** is the combination of Mendel’s genetics and Darwinism. It maintains that mutation provides the source of genetic variations, and natural selection acts on these mutations and determines the genetic makeup of populations in the process of gene substitution.

- **Selectionism:** is a branch of neo-Darwinism that believe natural selection is the only force to drive the evolutionary process, while other factors such as mutations and random genetic drift were minor contributors at best.

- According to selectionists, gene substitutions occur as a consequence of the selection for advantageous mutations (**positive selection**) and deletion of deleterious mutations (**purifying selection/negative selection**).
Neutral theory of molecular evolution

- With more molecular data becoming available, it is realized that there is more genetic variability in natural populations such as the Drosophila and human populations than originally imagined, and the difference among orthologous genes are even more striking.

- To explain these observations, Kimura (1968) and King & Jukes (1969) postulated that the majority of molecular changes in evolution are due to the random fixation of neutral or nearly neutral mutations.
Neutral theory of molecular evolution

This hypothesis now is known as the neutral theory of molecular evolution. It believes that at the molecular level, the majority of evolutionary changes among species and much of variability within a species are caused by random genetic drift of mutant alleles that are selectively neutral or nearly so.

According to this theory, the following conclusions can be drawn:
1. Positive selection of advantageous alleles and balancing selection play a little role in molecular evolution, because such mutations are very rare in nature.

2. The frequency of (neutral) alleles is determined largely by stochastic rules, and the picture that we observed at any given time is merely a transient state representing a temporary frame from an ongoing dynamic process.

3. Polymorphic loci consist of alleles that are either on their way to fixation or on their way to extinction. Thus, substitution and polymorphism are two facets of the same phenomenon.
Both selectionists and neutralists agree that a lot mutations are deleterious, and are rapidly wiped out from the population by purifying selection.

The major debating point between them is the proportion of advantageous mutation among the non-deleterious mutations.

Selectionists believe that majority of non-deleterious mutations are advantageous, and evolution is driven by the fixation of these advantageous mutations.

By contrast, neutralists maintain that majority of non-deleterious mutations are neutral, and evolution is largely driven by the random fixation of these neutral mutations, instead of advantageous mutations.