Chapter 10

GENETIC DRIFT: EVOLUTION AT RANDOM

Purposes and goals are unique to human thought.

Evolution occurs without purpose or goal, by chance.

Natural phenomena do not have purpose.

Scientists use chance or randomness, to mean that when physical causes can result in any of several outcomes, scientists cannot predict what the outcome will be in any particular case.

Scientists may be able to specify the probability, and thus the frequency, to have one or another outcome.

Genetic drift and natural selection are the two most important causes of allele substitution – evolutionary change – in populations.

Genetic drift, recombination and mutation provide the genetic material that natural selection can act on.

A sequence of random events in succession can change the frequency of an allele: this is called random genetic drift.

- Allele frequency is changed from one generation to the next: allele fluctuation.
- The genes in one generation do not end up in the same ratio in the next generation.
- This is a case of non-adaptive evolution: genetic drift does not necessarily results in adaptation.

We should not assume that a characteristic or a difference between populations and species is adaptive or has evolved by natural selection unless there is evidence for this conclusion.

THE THEORY OF GENETIC DRIFT

GENETIC DRIFT AS SAMPLING ERROR

The genes included in any generation, whether in newly formed zygotes or in offspring that survive to reproduce, are a sample of the genes carried by the previous generation.

Any sample is subject to random variation or sampling error.

The proportions of different kinds of items in a sample are likely to differ, by chance, from the proportions in the set of items from which the sample is drawn.

Over long periods of time, variation is more easily maintained in large populations than in small ones.

Random genetic drift reduces variation and leads eventually to the random fixation of one allele and the loss of other, unless other processes such as gene flow or mutation counter it.

When an allele reaches a frequency of 0.1 or 100% is said to be fixed. That population becomes monomorphic for that allele.

COALESCEENCE
Not all members of a generation leave equal number of descendants. Some do not leave any.

As time goes on, more and more alleles of a particular gene become extinct, so that the population consists of descendants of fewer and fewer of the original gene copies.

Eventually all gene copies in the population are descended form a single ancestral gene copy.

The genes in the present population have coalesced back to a single common ancestor.

The smaller the population, the more rapidly all gene copies coalesce back to a single ancestral copy.

A population will eventually become **monomorphic** for one allele or the other, and that the probability that one allele will be fixed rather than the other, is equal to the initial frequency of that allele.

The probability that an allele becomes fixed is equal to its frequency.

- \( A_1 = 0.8 \), the probability of \( A_1 \) to become fixed is 80%.
- \( A_2 = 0.2 \), the probability of \( A_2 \) to become fixed is 20%.

If a newly arisen mutation is represented by one copy in a diploid population of \( N \) individuals, the probability that it will increase fixation is \( 1/2N \).

“In genetics, **coalescent theory** states that all genes or alleles in a given population are ultimately inherited from a single ancestor shared by all members of the population, known as the most recent common ancestor. If the inheritance relationships are written in the form of a phylogenetic tree, termed a gene genealogy, the gene or allele of interest is said to undergo coalescence to the common ancestor..."

Coalescent theory seeks to reconstruct the ancestral relationship of individuals and is therefore of great utility in reconstructing the phylogenetic relationships of species based on information at the molecular level.”

http://en.wikipedia.org/wiki/Coalescent_theory

This does not mean that the population of a single common ancestor had only one member at that time.

Different alleles are fixed by chance in different populations.

This model does not include the evolution of adaptive traits, those that evolve by natural selection.

**RANDOM FLUCTUATIONS IN ALLELE FREQUENCIES**

Small independent populations of a species are called **demes**, and the grouping of all these population is a **metapopulation**.

An allele is selectively neutral if selection neither favors it or not.

Neutral alleles may change by mutation over time allowing the calculations of the molecular clock.

Example of **random fluctuation** of an allele:

A deme has an allele frequency of 0.5; in the next generation the frequency may be 0.47; in the following generation it will change again from 0.47 to some other higher or lower value. This process of random
fluctuation continues over time. Since no stabilizing force returns the allele to 0.5, the frequency will wander (drift) to either 0 or 1: \textit{the allele is either extinct or fixed}.

Once an allele has reached either 0 or 1, it cannot change unless another allele is introduced into the population, either by mutation or by gene flow from another population.

The allele frequency may increase in some demes of the metapopulation and decrease in some others.

The variance in allele frequency among the demes continues to increase from generation to generation.

The number of demes fixed for on or another allele continues to increase, until all demes have become fixed.

Thus, demes that initially are genetically identical evolve by chance to have different genetic constitutions. The assumption here is that the alleles are neutral.

\section*{EVOLUTION BY GENETIC DRIFT}

The following points are some of the most important aspects of evolution by genetic drift.

1. Allele or haplotype frequency fluctuates at random within a population, and eventually one or the other becomes fixed.
2. Genetic variation at a locus declines and eventually is lost. The frequency of heterozygotes declines.
3. At any time, an allele’s probability of fixation equals its frequency at that time, and is not affected or predicted by its previous history of change in frequency.
4. Populations with the same initial allele frequency diverge.
5. An allele just arisen by mutation and represented by only one among the 2N copies in the population, its frequency is \( p = 1/2N \) …and this is the likelihood of reaching a frequency of 1.
6. Evolution by genetic drift proceeds faster in small than in large populations.
7. Since the allele frequency in each deme changes, eventually becoming 0 or 1, the frequency of heterozygotes declines to zero in each deme and in the metapopulation as a whole.
8. Mildly advantageous or disadvantageous alleles may increase in frequency by genetic drift, and may even be fixed. Fitness, then, may decline if the fixed is a disadvantageous allele. This can be viewed as a sort of inbreeding depression.
9. The frequencies of two alleles at linked loci will fluctuate by genetic drift and one linked combination may increase to a frequency greater than expected creating a random association of alleles, that is, linkage disequilibrium. Recombination breaks down linkage disequilibrium.

\section*{Effective population size}

In many populations the number counted - \textbf{Census Size} - may be greater than the number that actually contributes genes to the next generation.

- Elephant seals; wolves.

The rate of genetic drift of allele frequencies, and of loss of heterozygosity, will be greater than expected from the population’s census size.
The **effective size** \( (N_e) \) of an actual population is the number of individuals in an ideal population (in which every adult reproduces) in which the rate of genetic drift (measured by the decline of heterozygosity) is the same as it would be in the actual population.

- The effective population size is the number of individuals in a population that contribute to the next generation.
- The effective population size is actually smaller than the census or absolute population.

**Reasons:**

1. Variation in the number of progeny produced by females, males or both reduces \( N_e \).
2. A sex ratio different from 1:1 lowers the effective population size.
3. Natural selection can lower \( N_e \), e.g. some individuals contribute more gene copies than others.
4. Generation overlap or inbreeding; when offspring mate with parent generation, the pair carries identical copies of the same genes, the number of genes propagated is reduced.
5. Fluctuation in population size. Smaller populations are more affected than larger populations.

The equilibrium of genetic variation is set by the balance between the gain of variation attributed to mutation and the loss attributed to genetic drift.

**Founder effect**

**Population bottlenecks** occur when a population is reduced for at least one generation.

“For Northern elephant seals have reduced genetic variation probably because of a population bottleneck humans inflicted on them in the 1890s. Hunting reduced their population size to as few as 20 individuals at the end of the 19th century. Their population has since rebounded to over 30,000—but their genes still carry the marks of this bottleneck: they have much less genetic variation than a population of southern elephant seals that was not so intensely hunted.”

http://evolution.berkeley.edu/evosite/evo101/IIID3Bottlenecks.shtml

The genetic drift that occurs when few colonists establish a new population is called **founder effect**.

- The **founder effect** was defined by Ernst Mayr in 1963.
- The founders carry only a small fraction of the genes found in the original population. There is a loss of genetic variation.
- The new population will be genetically different from the parent population: the frequency of genes will be different.
- In some cases, it may lead to speciation.

“For example, the Afrikaner population of Dutch settlers in South Africa is descended mainly from a few colonists. Today, the Afrikaner population has an unusually high frequency of the gene that causes Huntington’s disease, because those original Dutch colonists just happened to carry that gene with unusually high frequency. This effect is easy to recognize in genetic diseases, but of course, the frequencies of all sorts of genes are affected by founder events.”

http://evolution.berkeley.edu/evosite/evo101/IIID3Bottlenecks.shtml

**GENETIC DRIFT IN REAL POPULATIONS**

**Laboratory populations**

Experiment by Peter Buri (1956)
• Started 107 experimental populations of Drosophila melanogaster each with 8 males and 8 females.
• All heterozygous for two alleles \( bw \) and \( bw^{75} \) for eye color by which all three genotypes are recognized.
• Propagated each population for 19 generations.
• By generation 19, 30 populations had lost the \( bw^{75} \) allele, and 28 had become fixed for it.
• Among the unfixed populations, intermediate allele frequencies were quite evenly distributed.
• The results matched those expected by the genetic drift theory.

Other experiments conducted by McCommas and Bryant (1990) with houseflies also supports the mathematical predictions of the genetic drift theory.

• Four populations of houseflies were established at each three bottleneck sizes: 1, 4 and 16 pairs.
• Each population rapidly increased to 1000 individuals, after which the populations were reduced to same bottleneck size.
• The procedure was repeated five times.
• The average heterozygosity decline steadily after each bottleneck.
• The smaller the bottlenecks were, the more rapidly it declined.

Natural populations

Theories tell us what patterns to expect in the genetic features of natural populations.

We infer the causes of evolution by interpreting these patterns.

Patterns of molecular genetic variation in natural populations often correspond to what we would expect if the loci were affected by genetic drift.

Fixation of deleterious alleles can reduce survival and reproduction, increasing the risk of extinction.

**THE NEUTRAL THEORY OF MOLECULAR EVOLUTION**

We saw in chapter 9 that a high proportion of enzyme loci are polymorphic. Lewontin and Hubby argue that natural selection could not actively maintain so much genetic variation, and suggested that much of it might be selectively neutral.

**Neutral alleles** are not subject to natural selection because they do not affect or have very little effect on fitness.

“The neutral theory of molecular evolution is that most evolutionary change at the molecular level is driven by random drift rather than natural selection. The neutral theory does not suggest that random drift explains all evolutionary change: natural selection is still needed to explain adaptation. However, the neutral theory states that evolution at the level of the DNA and proteins, but not of morphology, is dominated by random processes; most evolution at the molecular level would then be non-adaptive.

The neutral theory can be contrasted with the idea that almost all molecular evolution has been driven by natural selection.”

Source: [http://www.blackwellpublishing.com/ridley/tutorials/Molecular_evolution_and_neutral_theory2.asp](http://www.blackwellpublishing.com/ridley/tutorials/Molecular_evolution_and_neutral_theory2.asp)

• Proposed by Motoo Kimura in the late 1960s and early 1970s.
- Small minority of mutations in DNA or protein sequences is advantageous and become fixed by natural selection.
- Many mutations are disadvantageous and become eliminated by natural selection.
- The great majority of mutations that are fixed do not affect fitness – they are neutral.
  - Silent or synonymous mutations.
- Neutral mutations are fixed by genetic drift because natural selection does not act on it.


The neutral theory of molecular evolution also proposes that:

- Evolutionary substitutions at the molecular level proceed at a roughly constant rate,
- So that the degree of sequence difference between species can serve as a molecular clock
- It is possible to determine the divergence time of species.

“Recent research suggests that many mutations in non-coding DNA do have slight deleterious effects.”[15][16]


**PRINCIPLES OF THE NEUTRAL THEORY**

Only a fraction of the mutations that occur is effectively neutral.

The neutral mutation rate, $\mu_0$, will be less than the total mutation rate, $\mu_T$.

Effectively neutral means that the mutant allele is so similar to the original one that in its effect on survival and reproduction that changes in its frequency are governed by genetic drift alone, not by natural selection.

If the mutation affects fitness to some slight extent then natural selection and genetic drift will operate simultaneously on the frequency of the mutant allele.

The majority of the mutations will be deleterious if the protein encoded by the gene cannot be altered without seriously affecting its function and the neutral mutation rate will be much lower.

This locus is said to have many functional constraints.

Highest mutation rate will be at the third-base-pair position within regions that code for proteins, and lowest at the second-base-pair position because these places have the highest and lowest redundancy.

- UUA and UUG both codons code for leucine

DNA sequences that are not transcribed and have no known function such introns and pseudogenes would have less or no constraints, and therefore, the neutral mutation rate will be the highest.
Applying the principles of the neutral theory, the following can be calculated knowing the size of the effective population, \( N_e \), and the rate of mutation per gamete per generation, \( u_0 \):

- Average number of new mutations
- Probability that a mutation will be fixed by genetic drift.
- Number of neutral mutations that arise in a given generation and will someday be fixed.

Therefore, the same number of neutral mutations should be fixed in every generation: the rate of fixation of mutations is theoretically constant, and equals the neutral mutation rate.

- The neutral theory predicts that over long spans of time, substitutions will occur at an approximately constant rate for a given gene providing a **basis for the molecular clock**.

From the above data, the number of generations that have passed since the divergence from a common ancestor of two species can be estimated.

As time passes, multiple or **successive substitutions** in the same locus causes the observed number of differences between species to be less than the number of substitutions that have occurred.

- \( A \to C \to T \to A \); it will appear that \( A \) has not mutated when in fact it has mutated three times.

**The Debate**

The debate is not about whether natural selection operates or not.

1) Some evolutionary biologists prefer the neutral theory to be the “null hypothesis” to explain genetic variation within species or populations.

Adaptation and natural selection should be the preferred explanation only if genetic drift cannot explain the data.

The neutral theory has replaced natural selection with randomness in most cases of mutation.

2) There is so much evidence that natural selection has shaped species’ characteristics, that selection should be the explanation of choice.

The neutral theory should be the preferred explanation only if natural selection cannot explain the data.

What percentage of the alleles is “neutral” versus those that are “non-neutral”?

The neutral theory states that:

- The majority of changes at the molecular level, DNA mutations, are neutral and do not affect fitness.
- Synonymous mutations and mutations in genes with little or no constraints will be fixed more rapidly than those with functional constraints that affect important functions.
- Genetic drift is the major driving force and natural selection is secondary in explaining genetic differences within populations.
• Natural selection should be used as the explanation for genetic differences only when genetic drift cannot explain the differences.

**Variation within and among species.**

Various kinds of DNA sequences or base pairs differ in their rate of neutral mutation because of differences in constraints and other factors.

The sequences of DNA or sites that differ more among related species should also display greater level of variation within the species.

There should be a positive correlation between the heterozygosity at a locus and its rate of evolution.

The fraction of mutations that are neutral is higher for synonymous than for non-synonymous (amino acid-replacing) nucleotide substitutions.

A study by McDonald and Kreitman (1991) showed that 5% of the polymorphisms, but fully 29% of the substitutions that distinguish species, are replacement changes.

They concluded that amino acid replacing substitutions is an adaptive process governed by natural selection.

If most replacement substitutions (non-synonymous) are advantageous rather than neutral, they will increase in frequency and be fixed more rapidly than by genetic drift alone.

Replacement substitutions will spend less time in the polymorphic state than selectively neutral synonymous substitutions do.

They will contribute less to polymorphic variation within species.

“PROTEINS often differ in amino-acid sequence across species. This difference has evolved by the accumulation of neutral mutations by random drift, the fixation of adaptive mutations by selection, or a mixture of the two. Here we propose a simple statistical test of the neutral protein evolution hypothesis based on a comparison of the number of amino-acid replacement substitutions to synonymous substitutions in the coding region of a locus. If the observed substitutions are neutral, the ratio of replacement to synonymous fixed differences between species should be the same as the ratio of replacement to synonymous polymorphisms within species. DNA sequence data on the Adh locus (encoding alcohol dehydrogenase, EC 1.1.1.1) in three species in the Drosophila melanogaster species subgroup do not fit this expectation; instead, there are more fixed replacement differences between species than expected. We suggest that these excess replacement substitutions result from adaptive fixation of selectively advantageous mutations.”


“...data from Drosophila simulans and D. yakuba. We estimate that 45% of all amino-acid substitutions have been fixed by natural selection, and that on average one adaptive substitution occurs every 45 years in these species.”


**Do comparisons among species support the neutral theory?**

Sequencing DNA has provided data on the rate of molecular evolution.

The rate of nucleotide or amino acid substitution can be estimated from the number of sequence differences among species in two ways.
1. An absolute rate can be estimated by a calibration based on fossil evidence of the time since two or more taxa diverged from their common ancestor.

2. The relative rates of evolution among different lineages can be estimated simply from the number of differences that have accumulated in each member of a monophyletic group, relative to an out group.

Based on the data provided by DNA sequencing, most DNA sequence evolution has been neutral.

1. Rate of synonymous substitutions is generally greater than the rate of replacement substitutions, that is, substitutions occur more frequently at the third-base position in codons and least frequently in second-base positions

2. Rates of substitutions are higher in introns than in coding regions of the same gene, and even higher in pseudogenes.
   - Pseudogenes are non-functional genes related in sequence to functional genes.

3. Some genes evolve more slowly than other, e.g. histone genes.
   - Genes that evolve most slowly are those thought to be most strongly constrained by their precise function.

Proteins interact to form complexes in order to carry out their biological function, e.g. hemoglobin is a complex of four proteins.

Some of these interactions are more permanent than others that are temporary.

Fraser and colleagues found an inverse relationship between the rate of amino acid substitutions and the number of proteins with which it interacts.

They also found that proteins in the same complex evolve at more similar rates than expected by chance, suggesting that interacting proteins coevolve.

The rate of evolution is greatest at DNA positions that, when altered, are least likely to affect function, and therefore least likely to alter the organism’s fitness.

Support for the neutral theory’s prediction that the rate of evolution should be constant among phyletic lineages is equivocal.

- Some rates have been constant and others not. See fig. 10.16.
- Mitochondrial DNA sequences have evolved more slowly in turtles than in other vertebrates.
- The rate of molecular evolution appears to be more nearly constant for synonymous than for nonsynonymous substitutions, as measured by differences among species.

**GENE FLOW AND GENETIC DRIFT**

**Gene flow or gene migration** is the transfer of alleles from one population to another.

The rate at which populations drift toward fixation of one allele or another is **inversely proportional to the effective population size**. The smaller the population, the faster the fixation rate.

- In smaller populations genes become fixed sooner than in larger populations.
Gene flow counteracts gene fixation.

The mathematical model (page 272) shows that a little gene flow keeps all the demes fairly similar in allele frequency and heterozygosity remains high.

Genetic drift and gene flow affect all loci the same way, whereas natural selection affects different loci more or less independently.

If each of a number of polymorphic loci yields about the same value of gene frequency, it is likely that selection is not strong.

**Gene trees and population history**

The genealogical history of genes in populations is the basis of *Coalescent Theory*.

Because gene lineages within a population become extinct by chance over the course of time, all gene copies in a population today are descended from one gene copy that existed at some time in the past.

The smaller the effective size of a population, the more rapidly genetic drift transpires.

The existing gene copies in a small population must stem from a more recent common ancestor than the gene copies in a large population.

It takes longer for the present genes in the larger population to coalesce to their common ancestor.

In a diploid population, the common ancestor of a random pair of gene copies occurred $4N_e$ generations ago, where $N_e$ is the size of the effective population, and $t_{CA}$ is the average time back to the common ancestor.

$$t_{CA} = 4N_e$$

**THE ORIGIN OF MODERN HOMO SAPIENS REVISITED**

Homo erectus migrated from Africa to Europe and Asia about 1.8 million years ago.

H. erectus evolved there into Homo neanderthalensis and “archaic Homo sapiens”

Mitochondrial DNA provides the evidence that supports the “replacement” or “out-of Africa” hypothesis.

Studies of mtDNA have concluded that all human mitochondrial genes, both African and non-African, are descended from a common ancestral gene that existed at $t_{CA} = 156-250$ Kya.

- Similar conclusions were reached in analyses of Y chromosome sequences and autosomal microsatellite loci.

This does not mean that the human population at that time consisted of one man and one woman; it means only that all other mitochondria and Y chromosomes in the population at that time have failed to leave descendants.
There is considerable evidence that modern Homo sapiens, a separate species, evolved in Africa and spread from there about 100,000 years ago, replacing “archaic Homo sapiens” without interbreeding with them to any great extent.

The conclusion of these studies is that the world human population is descended from a breeding population of 4,600 to 11,200 individuals.

A study conducted by Garrigan and colleagues (2007) concluded that the ancestral Eurasian population was founded only about 40 Kya. These estimates are much more recent than they would be if the major regional populations were descended from the archaic Homo populations that inhabited those regions.

There are several mtDNA lineages in Africa.

There are far fewer nucleotide differences among non-African haplotypes, on average, than among African haplotypes.

Similar conclusions were reached in analyses of Y chromosome sequences and autosomal microsatellite loci.

- Supporting evidence synthesizes new genetic, archaeological, climatic, and linguistic discoveries.
- Genetic diversity amongst modern human populations is greatest in the African continent.
- There are 15 surviving mtDNA lineages in Africa which date back more than 80,000 years, compared with only one for the rest of the world.
- This population probably consisted of a breeding population between 4,600 and 11,200 individuals.

Three distinct African populations can be determined genetically on the basis of the L0 mutation in mtDNA, aligning with ethnic, linguistic and now archaeological evidence.

1. mtDNA L1. Y-chromosome haplogroup A = M91. The Khoisan speakers - the modern Hadza of Tanzania and the Bushmen (San) and Khoikhoi (once called "Hottentot") populations.

2. mtDNA L2 (85-95,000 yrs ago), Y chr B = M60 (began 50-60,000 yrs ago): This population adopted cultural innovations suitable for rainforest and woodland zones.
   - The West African genomes of the Niger-Congo and Nilo-Saharan speaking peoples - led to the evolution of the Ivory and West Africa and modern Bantu populations.
   - Also sharing the Y chr B , mtDNA L2 are the smaller groups surviving in the Congo rainforests, the Mbuti and Bayaka pygmy peoples of the Ituri and other Congo rainforests.

3. mtDNA L3: Human populations similar to those found at Ngaloba and Laetoli in East Africa, and dated to 59-69,000 years before the present.
   - All other groups of human probably descended from this group.