February 8, 2005

Bio 107/207
Winter 2005
Lecture 11
Mutation and transposable elements

- the term "**mutation**" has an interesting history.
- as far back as the 17th century, it was used to describe any drastic change in an organism's form, such as that found in the fossil record.
- in the early part of the 20th century, the term mutation was used by Hugo deVries to refer to the origin of new species.
- deVries' "**Mutation Theory**", in which species were formed by large jumps (saltations), was more popular than Darwin's theory of evolution for a period of time.
- when deVries' theory fell into disrepute, the term "mutation" was rescued from a sure death by Thomas Hunt Morgan (one of the founders of the science of genetics).
- Morgan used it to refer to newly arisen aberrations in phenotype (such as the appearance of white-eyed Drosophila) that obeyed Mendelian patterns of inheritance.
- this nomenclature has been maintained up to the present day.

- mutations can occur at different levels – from single nucleotide positions in DNA to entire genomes.

Point mutations

- the simplest types of mutations occur at single base positions of DNA.
- four categories of point mutations are recognized:

1. transitions: A to G, C to T
2. tranversions: T to A, C to G
3. insertions
4. deletions

- within coding regions point mutations can involve **silent** (or **synonymous**) or **replacement changes** (or **nonsynonymous**).
- insertions/deletions occurring within coding regions of DNA give rise to **frameshift mutations** whereby the correct reading frame is disrupted.
- when we monitor mutational changes within genes over evolutionary time, transitions greatly outnumber transversions (because they are easier mistakes to make) and silent changes greatly outnumber replacement changes (typically).

2. "Copy-number" mutations

- these mutations change the numbers of genetic elements.
- included in this category are **gene duplication** events where a single gene is duplicated.
- this duplication process typically results in the new gene copy being close to the progenitor copy.

- over time, this process may lead to the development of **gene families** where a number of copies are physically clustered in chromosomes.
- good examples are the alpha and beta globin gene families (on chromosomes 16 and 11 in humans, respectively).
- the mechanism believed responsible is **unequal crossing over**.
- once a couple of copies of a gene are closely placed, it may be relatively easy to generate more.

- another major source of mutations that can have important effects at the phenotypic level involves **transposable elements (TE's)**.
- TE's are also called jumping genes because they can move about the genome, copying themselves and inserting into different places (a process called replicative transposition).
- three major classes of TE's are recognized.

1. insertion sequences (700-2600 bp)
2. transposons (2500-7000 bp)
3. retroelements

- TE's have a multitude of effects including altering gene expression, increasing mutation rates, and causing chromosomal rearrangements.

3. Chromosomal mutations

- the two major types of chromosomal mutations are **inversions** and **translocations**.
- **inversions** occur when a piece of a chromosomal flips around and becomes integrated in the opposite direction.
- inversions are particularly common in insects for reasons that are not well understood.
- they form an important type of polymorphism in many insect species, notably Drosophila, where they have been studied intensively.
- inversions accomplish something important: they lock up the block of genes as a whole unit and suppress recombination among the block.
- (this occurs because when recombination occurs in an inversion heterozygote, the crossover products contain duplications and deletions.)
- as a result of this suppression of recombination, the genes locked within an inversion can coevolve to function optimally as a concerted group giving rise to what are called **coadapted gene complexes**.
- **translocations** occur when a piece of chromosomes breaks off and moves to a new region.
- these may have important effects by modifying the level of transcription of genes present in the translocated segment, especially near the breakpoint.
- some translocations in humans have been linked to certain cancers and other diseases.
- finally mutations affecting the **karyotype** of a species can occur.
- Robertsonian fusions and fissions are fairly common and can have major effects on speciation.

4. Genome mutations

- the "highest" level that mutations can occur are those affecting the entire genome.
- such mutations are called **polyploidization events.**

- for example, diploid (2N) species can give rise to tetraploid (4N) species.
- tetraploids can produce octoploids, etc.
- as with less dramatic changes in chromosome number polyploidization events can have important effects in creating new species.
- polyploidy has played a major role in the evolution of plants.
- about half of angiosperms (flowering plants) are polyploids and an even higher number of grasses are too.

Distribution of fitness effects of new mutations

- what is the distribution of fitness effects of new mutations?
- the vast majority of mutations are harmful or deleterious.
- a substantial number have effects on fitness that are close to zero - these are called neutral or nearly neutral mutations.
- a small proportion are advantageous.
- how small this portion is remains unclear but they are responsible for driving all adaptation.

Mutation as a random process

- mutation is a random process.
- it is important to understand what this statement means and what it doesn't mean.
- it does not mean that all conceivable mutations are likely to occur.
- developmental "constraints" might exist that prevent a large class of mutations from ever being realized at the phenotypic level.

- what we mean when we say mutation is random is that the probability that a certain mutation will occur is not influenced by whether an organism is in an environment in which that mutation would be advantageous.
- that is **the environment does not induce adaptive mutations**.
- the idea that mutations are somehow "directed" towards benefiting their carriers is an old one.
- this would indeed introduce a Lamarckian element to evolution for organisms could then acquire adaptive hereditary characteristics in direct response to their environments.
- there is a small group of evolutionary biologists who still argue that in some instances adaptive mutations can be affected by the environment.
- this is not widely accepted and it is difficult to conceive of a mechanism by which this can occur.
- how could the environment ever direct the mutation process by dictating just base pair change be made?

The fate of new mutations

- the fate of the vast majority of new mutations is that they will be lost in first or subsequent generations following their appearance.
- consider a neutral mutation.
- if the mutation is unique, it will occur in a single heterozygote in the population at a frequency of 1/2N.

- if we assume a Poisson distribution of family size, it is possible to predict the probability that the mutation will be lost in one generation to be 0.368.
- in the second generation, the probability of its loss is 0.532.
- by five generations, the probability increases to 0.732.

- what about a selected allele?
- surprisingly, the fate of new advantageous mutations is not that much better than neutral alleles!
- if the selection coefficient is 0.01, then after one generation probability of the selected allele's loss is improved to only 0.364.
- after two generations, its probability of loss is improved to only 0.526 and after 5 generations to only 0.725!
- in fact, the overall probability of fixation of a selected allele is only twice the selection coefficient – which in this case is 0.02.
- this shows that stochastic processes dominate the overall fate of even selected alleles and that natural populations lose the vast majority of beneficial mutations and thus never realize their benefits!

Allele frequency change caused by mutation

- let us assume there are two alleles at a locus, A_1 and A_2.
- let p be the frequency of A_1 and q be the frequency of A_2.
- let the A_1 allele be the normal "wild-type" and A_2 be a detrimental allele.
- let the "forward" mutation rate from A_1 to A_2 be u and the "reverse" mutation rate from A_2 to A_1 be v.

$$A_1 \underset{u}{\overset{v}{\leftrightarrows}} A_2$$

- it is generally assumed that forward mutation rates occur more frequently because it is easier to impair the function of a gene by randomly changing any given site than to restore the normal function.

- the change in the frequency of A_2 due to mutation is

$$\begin{aligned} \Delta q &= up - vq \\ &= u - q(u + v) \end{aligned}$$

- if no other forces at acting at the locus, there will be an equilibrium frequency established that depends on the rates of u and v:

$$q_e = u/(u + v)$$

- because mutation rates are generally very small, the rate of allele frequency change caused by mutation is very slow.
- let us assume that v is small relative to u (so it can be ignored).
- the change in the frequency of A1 after one generation of mutation is:

$$p_1 = p_0 - up_0$$

$$= (1 - u)p_0$$

- after t generations this relationship can be generalized to

$$p_t = (1 - u)^t p_0$$

- if $u = 10^{-5}$ and $p_0 = 0.50$, then this equation predicts that it would take 69,314 generations for the A_1 allele to reach 0.25.
- we can assess the effect of back mutation by including it in this equation:

$$p_1 = (1 - u)p_0 + v(1 - p_0)$$

- this can be generalized to

$$p_t = v/(u + v) + [p_0 - v/(u + v)] (1 - u - v)^t$$

- if we begin again with $p_0 = 0.50$ and let $u = 10^{-5}$ and $v = 10^{-6}$, then after 69,314 generations the frequency of A_1 is reduced to 0.282.
- this shows that a small rate of back mutation has a minor effect.
- however, the important point is that mutation is an extremely weak evolutionary force.

Mutation-selection balance

- let us now incorporate the effect of selection acting on deleterious alleles.
- assume first that the deleterious mutation is completely recessive.
- the change in allele frequency caused by selection is:

$$\Delta q_s = -sq^2p/(1 - sq^2)$$

- although selection is acting to reduce the frequency there is an increase every generation caused by mutation:

$$\Delta q_{mu} = up$$

- if we assume the rate of back mutation is low, we can ignore its effect.
- because the effects of selection and mutation are in opposite directions, there is some point where they balance out:

$$\Delta q_s + \Delta q_{mu} = 0$$

$$up = -sq^2p/(1 - sq^2)$$

- if q^2 is small, this equation becomes

$$up = -sq^2p$$

$$q^2 = u/s$$

$$q_e = (u/s)^{1/2}$$

- this point is referred to as **mutation-selection balance**.

- if there is inbreeding in the population, then this has an effect on the equilibrium:

$$q_e = u/fs$$

- this equation shows that the equilibrium frequency is greatly decreased compared to that in an outbreeding population.
- this occurs because inbreeding exposes deleterious alleles in homozygous state thus allowing selection to remove them from the population.
- note that inbreeding is not causing allele frequencies to change.
- rather, it is the action of purifying selection.

The effect of dominance

- if there is some dominance, then the equilibrium is greatly influenced.
- the degree of dominance at a locus is represented by the coefficient, h.

Genotype	A_1A_1	A_1A_2	A_2A_2
Fitness	1	1 – hs	1 – s

- different values of h can give rise to numerous outcomes.
- if h =1, then the A_2 allele is completely dominant.
- if h = 0, then A_2 is completely recessive.
- if h = 0. 5 then the alleles behave in additive fashion.
- if h is negative, there is overdominance!
- if h is between 0.5 and 1, there is some degree of dominance.

- if h is close to 1 and q_e is small, then the equilibrium allele frequency is:

$$q_e = u/hs$$

- this shows what our intuition predicts that dominance of a deleterious allele results in a substantially lower equilibrium frequency.

Mutational load

- the persistence of deleterious alleles at mutation-selection balance imposes a genetic load on a population.
- the mean population fitness would be greater if these deleterious alleles were not present.
- however, for outbreeding species, it is impossible to purge these alleles.
- and, if they are removed, they are continuously re-introduced by recurrent mutation.
- for a recessive mutation, the load, L is

$$L = sq^2$$

- since $u = sq^2$ this means that the load is roughly equal to the mutation rate:

$$L \approx u$$

- in the case of complete dominance, there are very few A2A2 homozygotes, thus the load becomes:

$$L = 2pqhs$$

- using the equation 7.5d in the textbook, the load then becomes

$$L \approx 2u$$

- therefore, the mutational load of a population is between the mutation rate (u) and twice the mutation (2u) depending on the degree of dominance.
- note that the mutation load is independent of the intensity of selection!

Estimation of mutation rates

- the estimation of mutation rates are difficult because mutations are rare events.
- mutation rates can be estimated directly from examining the progeny produced from known parental genotypes.
- the textbook describes some examples of studies using this direct count method.
- if the genotypes of the parents are not known, then an indirect method can be used.
- these indirect methods are most tractable for dominant mutations but the approach can also work for recessives.
- for most protein coding genes, mutation rates are typically in the range of 10^{-5} to 10^{-6} (per gamete per individual).
- for some highly variable markers like microsatellites, mutation rates are much higher – usually in the range of 10^{-2} to 10^{-3}.
- many studies have confirmed that the mutation rate in males is much greater than females.
- this observation is consistent with the much larger number of cell divisions that occur during spermatogenesis over the lifetime of a male.

- when mutation rates are compared among species, there is commonly a large difference in the mutation rate per sexual generation.
- if one corrects for the differences in genome size and differences in the number of cell replications per sexual generation there is little variation among fruit flies, *C. elegans*, or humans!
- mutation rates may, however, differ between closely-related species indicating that it is a trait that may evolve.
- one might predict that is variable or unpredictable environments, there might be an advantage to a higher mutation rate.

Transposable elements

- a transposable element (TE) is a DNA sequence that can change its location in the genome.
- transposable elements can give rise to chromosomal rearrangements, inactivate genes (if they insert within a transcribed gene), or modify the expression of a gene via their effects on gene regulation.
- the process of transposition requires an enzyme called transposase which is usually encoded in the DNA sequence of the TE.
- several models of the population dynamics of transposable elements have been constructed.
- they typically incorporate the following features:

1. A rate of infection where genomes lacking the TE become infected (u).
2. A rate of transposition in which copy number increases in a genome (T).
3. Natural selection acts to remove TEs from the genome because they come at a cost to an organism's fitness (cost S).

- these models have been found to result in some equilibrium distribution of elements.
- here is one example of IS30 elements in 71 natural isolates of *E. coli* from a study by Sawyer et al. (1987).

No. of IS30 Elements	Expected No. of strains	Observed No. of strains
0	35.5	36
1	17.8	16
2	8.9	13
3	4.4	2
4	2.2	2
≥ 5	2.2	2