Neutral theory 3:

Rates and patterns of molecular evolution

Predictions of the neutral theory

1. Within species variation is correlated with divergence between species.

2. Evolutionary rate is inversely related functional constraint.

2. Base composition at neutral sites reflects mutational equilibrium.

4. A molecular clock.

Neutral theory is “the” rigid null hypothesis for molecular evolution
1. Polymorphism and divergence are correlated

Neutral theory is a bridge between microevolution and macroevolution

Neutral population polymorphism within species is correlated with neutral divergence between species

1. Variation within and among species: polymorphism & divergence

This is one place where the genetic code is relevant:
1. Synonymous (S)
2. Non-synonymous (NS)

Neutrality and selection have different impacts on polymorphism:
1. Neutrality: NS residence times determined by \( N_s \)
2. Selection: NS residence times reduced by natural selection

Let's look at the ratio NS:S [ratio of counts]
1. Variation within and among species: polymorphism & divergence

Comparison of the ratio of synonymous and nonsynonymous polymorphism within species to divergence between species. Neutral theory suggests that the fraction of variation that is nonsynonymous within species should be the same as between species.

<table>
<thead>
<tr>
<th>Species 1</th>
<th>Species 2</th>
<th>Species 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polymorphic</td>
<td>12:4</td>
<td>6:2</td>
</tr>
<tr>
<td>Synonymous (S)</td>
<td>17:6</td>
<td>14:5</td>
</tr>
<tr>
<td>Non-synonymous (NS)</td>
<td>50</td>
<td>17</td>
</tr>
<tr>
<td>S:NS</td>
<td>3.1</td>
<td>2.9</td>
</tr>
</tbody>
</table>

Data are hypothetical. Ratios are tested by using a G-test on the counts of S and NS. These hypothetical data are not significant. If positive selection were acting, residence times for NS would be lower within species and polymorphic S:NS > fixed S:NS.

Genealogies within populations

Species level phylogenies

2. Rate of evolution is inversely related to functional constraint

Rate variation is well known:

- Fast genes (D-loop) verses slow genes (Histones)
- Introns verses exons
- Synonymous verse nonsynonymous sites

Neutral theory is consistent with such rate variation

- Asserts only that polymorphism is *selectively equivalent*
- Frequency of such polymorphism can change among genes, sites etc.
Note: two ways it is commonly measured

2. Rate of evolution is inversely related to functional constraint

Mean number of substitution per site at the three codon positions of the epsilon-globin gene of primates. Two measures are presented: (i) the average over all pairwise comparisons between genes; and (ii) the sum of the branch lengths of the epsilon globin gene tree.

Note: mean number of substitutions per site were computed in all cases by using the Jukes and Cantor (1969) correction.
2. Rate of evolution is inversely related to functional constraint

Kimura, 1968:

The neutral mutation rate per site is:

$$\mu_0 = \mu f_0$$

The neutral substitution rate per site is:

$$k = \mu f_0$$

The rate of evolution depends on the “size ($f_0$) of the selective sieve”

Kimura’s $f_0$ is the fraction of mutations that passes through the “sieve”
2. Rate of evolution is inversely related to functional constraint

Example: 3rd codon positions verses synonymous sites

Some changes at 3rd codon positions are NOT synonymous

Prediction:
1. $f_0$ for 3rd codon positions $< f_0$ for synonymous sites
2. rate 3rd codon positions $<$ rate for synonymous sites

The average substitution rate between primates and rodents is higher for synonymous sites as compared with third codon positions. The results are based on a sample of 82 nuclear genes.

Mean number of substitutions per site between primates and rodents is $r = t_0 + t_1$. The unit of time is $2 \times 80$ my; the time since primates and rodents shared a common ancestor.

This result is consistent with neutral theory given that $f_0$ is smaller for 3rd codon positions because some mutations at such site will be nonsynonymous.

2. Rate of evolution is inversely related to functional constraint

We can put sites into a wide variety of categories:

- 5’ and 3’ flanking regions
- 5’ and 3’ untranslated regions
- Introns
- Exons
- 3rd positions of 4-fold degenerate codons
- Nonsynonymous sites of a codon
- Functional domains
- **Pseudogenes**

Comparison of mean substitution rates in different parts of genes and pseudo-genes. Data is from Li et al. (1985). Substitution rate is the mean number of substitutions per site per 10^9 years. Rates are an average over 3000 mammalian genes.
2. Rate of evolution is inversely related to functional constraint

Sites subject to selection will have variable $f_0$ depending on the level of functional constraints acting on that site:

1. Nonsynonymous sites
2. Functional domains
3. Etc.

Note: many of the shaded sites are located in the heme pocket or at the interfaces between globins subunits, consistent with the notion that sites most critical to protein function evolve at the slowest rates.
2. Rate of evolution: differences among genes

Let's estimate the width of the selective sieve:

Under neutral theory:

- The synonymous substitution rate \( k_S \) is equal to the neutral mutation rate.
- The nonsynonymous substitution rate \( k_N \) measures the substitution rate for neutral amino acid changes.
- Thus the ratio of these rates \( k_N / k_S \) represents the fraction of amino acid mutations that are neutral: this is \( f_0 \) for amino acids.
- The fraction of amino acid mutations that are deleterious \( (1 - f_0) \) must be \( 1 - k_N / k_S \).

Let's take the Neuroleukin gene of primates as an example:

\[
\begin{align*}
    k_S &= 0.016 \\
    k_N &= 0.300
\end{align*}
\]

The fraction of amino acid changes that are neutral is \( 0.016 / 0.300 = 0.053 \), a small amount.
Hence the fraction of amino acid changes that are deleterious is \( 1 - 0.053 = 0.95 \)!
2. Rate of evolution: differences among genes

1. Nonsynonymous rates vary among genes due to differences in functional constraints

2. Synonymous rates vary due to differences in mutation rates [and in some cases weak selective constraints]

Best to estimate $k_S$ separately for each gene when trying to estimate $f_0$ and $f_D$. 
2. Rate of evolution: differences among genes

Distribution of nonsynonymous and synonymous substitution rates for 82 nuclear genes of primates.

Mean rate of nonsynonymous substitution:
0.045 / site / 80 million years

Mean rate of synonymous substitution:
0.201 / site / 80 million years

Data from Bielawski, Dunn, and Yang (2000) Genetics, 156:1299-1308.
Method: GY94 under ML

3. Patterns of mutation

Neutral theory: nucleotide frequencies at site free from selection will reflect mutational equilibrium.

For example:
- Pseudogenes
- 3’ flanking regions
- Synonymous sites
3. Patterns of mutation: human beta globin

Nucleotide frequencies in the human beta-globin gene differ among the three positions of the codon. Frequencies at positions 1 and 2 reflect selection acting on the protein product of the gene. Frequencies at position 3 reflect a strong influence of mutation pressure.

4. Molecular clock

\[ k = \mu \]

Neutral theory (1968) predicts that the rate of molecular evolution [substitution] should be approximately constant over time, where time is measured in generations.

Zukerkandl and Pauling (1965) noticed an approximately uniform rate of amino acid substitution, with time measured in years.

The notion of a molecular clock is somewhat controversial.
4. Molecular clock

Linear relation between mitochondrial substitution rate and time since common ancestor in teleost fishes

Linear relationship is expected under a uniform rate of substitution. Substitutions are the mean number of changes at first codon positions of all mitochondrial protein coding genes. Data were kindly provided by K. Dunn.

Note: many proteins exhibit constant rate in terms of absolute time in years

Some initial problems with neutral theory

1. Expected heterozygosity was larger than observed in natural populations

2. Some molecules evolved according to a per million year clock even though generation times were very different

3. There were more problems to follow, but will not consider them.
Nearly neutral theory

Ohta: What happens if some mutations are only mildly deleterious?

Nearly neutral theory

Slightly deleterious mutations:

- Small selective coefficients \( s \)
- When \( s \) is small, populations size plays important role:
  - Large \( N_e \): selection effective
  - Small \( N_e \): selection less effective [more deleterious alleles get fixed!]

Ohta and Kimura (1971): the slightly deleterious model of evolution
- more adjustments were to follow
Then fate of a beneficial recessive allele (A1) is not always predictable under the combined effects of directional selection and genetic drift. If there is no genetic drift (left: \( N_a = \infty \)), the fate of the recessive allele (A1) is always determined by selection. When there is drift (right: \( N_a < \infty \)) the fate of the recessive allele (A1) is not necessarily determined by selection; hence a deleterious allele can be fixed in a population.

Note that \( N_a > 1 \) does not guarantee that an allele is going to be fixed; it simply indicates that (as a long term average) the frequency that it is fixed will be greater than the frequency under genetic drift alone.
Nearly neutral theory

Slightly beneficial models were incorporated later

<table>
<thead>
<tr>
<th>Large population size: (selection very effective)</th>
<th>Small population size: (selection a little less effective)</th>
</tr>
</thead>
<tbody>
<tr>
<td>neutral mutations</td>
<td>neutral mutations</td>
</tr>
<tr>
<td>slightly beneficial mutations</td>
<td>slightly beneficial mutations</td>
</tr>
<tr>
<td>FIXED</td>
<td>FIXED</td>
</tr>
</tbody>
</table>

Mildly deleterious + slightly beneficial = “Nearly neutral model”

The strictly neutral model was extended to accommodate nearly neutral mutations

Fraction of “neutral” mutations \( f_0 \) changes with \( N_e \)

Note that \( N_e \) changes over time!
Nearly neutral theory reconciles some observations with theory

1. Predicts lower natural levels of heterozygosity

2. Possible reconciliation of predicted rate constancy in generations with rate constancy in years.

Neutrality depends on environmental conditions

Strictly neutral and nearly neutral models:
- distribution of fitness effects of new mutations changes according to environment

Genetic environment and physical environment change:
- 3D space of protein reflect a genetic as well as physical environment
- neutral substitutions, recombination, LGT, etc. change the genetic environment
- physical environment changes daily, weekly, seasonally, yearly….
Success of neutral theory: “the null model”

Rob Stainer (1970) at the general meeting of the society for general Microbiology commented that evolutionary studies are
“a relatively harmless habit, like eating peanuts”.

Neutral theory provides the foundation for the science of molecular evolution

Success of neutral theory: James F. Crow (1985)

1. The theory provides the best explanation for the dramatic differences in the rates and patterns of evolution in molecules as compared with morphology.

2. The neutral theory provides a common framework for understanding the dramatic differences among genes, codon positions, introns, and pseudogenes.

3. The neutral theory correctly predicts the differences in rates among molecular datasets as well as the similarity of substitution rates between the so-called “living fossil” organisms and the most rapidly changing species.

4. The neutral theory has stimulated theoretical studies as well as studies of natural variation in a framework based on a rigid null hypothesis.