Neutral theory of molecular evolution

Core course: ZOOL 3014
Semester: VI
Prof. Pranveer Singh

The neutral theory, formally known as the neutral theory of molecular evolution, was independently proposed by M. Kimura in 1968

And

J. L. King and T. H. Jukes in 1969

Kimura proposed that most genetic changes have been fixed by random drift rather than positive Darwinian selection

Kimura 1968 further states that the large amounts of fruit fly (Lewontin and Hubby 1966) and human (Harris 1966) genetic polymorphism discovered by using protein gel electrophoresis is consistent with the hypothesis that natural polymorphisms are largely neutral.

All populations have genetic variation that may or may not affect the fitness of the organism

Neutral theory of molecular evolution

The neutral theory of molecular evolution suggests that most of the genetic variation in populations is the result of mutation and genetic drift and not selection

According to this theory, if a population carries several different alleles of a particular gene, variation is neutral: having allele A or allele B does not affect your fitness

Neutral theory of molecular evolution asserts only that observed amino acid substitutions and polymorphisms are effectively neutral, not that the loci involved are unimportant or that allelic differences at those loci have no effect on fitness

The chief tenet of the theory is that most genetic differences between species and polymorphisms within species are selectively neutral and result from mutation and genetic drift

This view sharply contrasts that of neo-Darwinists, who maintain that most of these variations are adaptive

King and Jukes argue that synonymous substitutions, which do not alter protein sequence, are most likely neutral. But, Clarke and Richmond argue that synonymous codon usage is potentially selected to optimize translational efficiency

Evidence for the Neutral Theory

Zuckerkandl and Pauling 1965 proposes the concept of "molecular clock" based on the authors' observation that the aminoacid substitution rate per year for a protein is more or less constant across different evolutionary lineages

Because the rate of neutral substitution equals the rate of neutral mutation, neutral theory can explain the molecular-clock phenomenon if the neutral mutation rate is constant per year

Neutral theory predicts that the evolutionary rate of a gene or a site increases as its functional constraint reduces, because a reduced functional constraint means an increased fraction of neutral mutations

Kimura and Ohta 1974 shows that the functionally important and constrained histone H4 evolves much more slowly than the functionally relatively unimportant and unconstrained fibrinopeptides, consistent with the prediction of the neutral theory

Synonymous sites are functionally less constrained than nonsynonymous sites, the preponderance of substitutions per synonymous site in the evolution of proteincoding genes also supports the neutral theory

As a null hypothesis, neutrality applies to phenotypic evolution such as the evolution of morphological and physiological traits

Neutral theory also applies to phenotypic traits in general

Exceedingly rapid evolution of functionless genes known as pseudogenes, compared with functional genes, strongly supporting the neutral theory

Nei and Graur 1984 analyzes intraspecific protein polymorphisms measured by gel electrophoresis from seventyseven species. The authors report that the data are generally consistent with the prediction of neutral theory under population bottlenecks but are incompatible with the model of frequent overdominant selection

Arguments against neutral theory (NeutralistSelectionist Debate)

The substitution rate of a protein is far from constant and that this inconstancy was considered to reflect the action of positive selection

Mutation rate is commonly believed to be constant per generation, rather than per year. Thus, it has been argued that the observed molecular clock per year cannot be explained by the neutral theory

Synonymous substitutions were initially believed to be the best example of neutral changes, but Ikemura 1981 finds that codons with high cognatetRNA concentrations are used more often than other synonymous codons of the same amino acid, suggesting that synonymous substitutions are subject to natural selection in relation to translation

According to the neutral theory, the amount of genetic polymorphism in a population increases with the effective population size, but the lack of populations with very high protein polymorphisms, referred to as "invariance of heterozygosity" in Lewontin 1974 (cited under Origin of the Theory), poses a challenge to the neutral theory

The Nearly Neutral Theory

Kimura defined neutral mutations by $|2Ns| \ll 1$, where N is the effective population size and s is the selection coefficient

The nearly neutral theory was mainly developed by Ohta and she defines nearly neutral mutations by $|Ns| \sim 1$

Unlike neutral mutations, whose fate is independent of effective population size, the fate of nearly neutral mutations depends on the effective population size

The theory was originally proposed in Ohta 1972a to explain why the protein evolutionary rate is approximately constant per year, while the DNA evolutionary rate shows a generation time effect (i.e., higher rates for species with shorter generations)

The nearly neutral theory predicts a negative correlation between the protein evolutionary rate and population size

e.g. As confirmed in genomescale analysis, for example, by Rhesus Macaque Genome Sequencing and Analysis Consortium 2007

A comparison between conservative and radical nonsynonymous substitution rates in Zhang 2000 also supports the nearly neutral theory

Comparisons between freeliving bacteria and related endosymbiotic bacteria in molecular evolution rates in Moran 1996 strongly support the theory

Neutrality Tests

Divergence Data of DNA Sequences

Using divergence data compare the number of synonymous substitutions per synonymous site (d_S) with the corresponding number of nonsynonymous substitutions per nonsynonymous site (d_N)

Under the null hypothesis of neutrality,

$$d_S = d_N$$

However, if $d_N > d_S$, (Positive selection)
if $d_N < d_S$, , (Negative selection)

<u>Polymorphism Data of DNA Sequences</u>

Depend on frequencies of variants at polymorphic nucleotide sites

Additional information on the linkage phase among variant sites and score a haplotype as an allele

Patterns of linkage disequilibrium such as the extended haplotype homozygosity (EHH) test

Polymorphism and Divergence Data

The neutral theory can also be tested by comparing polymorphism and divergence data

Hudson, et al. 1987 proposes such a test by comparing the ratio of divergence and polymorphism between two loci

Neutrality is rejected (i.e., at least one locus is under purifying or positive selection) when this ratio is unequal between two loci

McDonald and Kreitman 1991 revises the test in Hudson, et al. 1987 by comparing synonymous sites and nonsynonymous sites of the same gene instead of comparing two loci

Under the assumption that synonymous sites are neutral, a rejection of the null hypothesis by the McDonaldKreitman test may indicate the action of position selection for or negative selection against nonsynonymous substitutions

When,

$$(D_N/P_N) > (D_S/P_S)$$

Positive selection for interspecific nonsynonymous differences is inferred

$$\text{If, } (D_N/P_N) <<< (D_S/P_S)$$

Negative selection against interspecific nonsynonymous differences is inferred

Where, (D_N) is interspecific nonsynonymous differences
(P_N) is intraspecific nonsynonymous polymorphisms
(D_S) is interspecific synonymous differences
(P_S) is intraspecific synonymous polymorphisms

Adaptive Molecular Evolution

Adaptive Protein Evolution Detected by Neutrality Tests from Studies of Individual Genes

Numerous tests of the hypothesis of neutral evolution have been conducted for DNA sequences. In a number of cases, neutrality is rejected in favor of adaptation driven by positive Darwinian selection

Hughes and Nei 1988 shows that (d_N) is significantly greater than (D_S) in the antigen recognition regions of the human major-histocompatibilitycomplex (MHC) genes, suggesting the action of positive selection, which is most likely related to the immune function of MHC

Numerous genes have since been shown to be subject to recurrent or episodic positive selection on nonsynonymous changes. Bestknown examples include the hemagglutinin gene in human influenza viruses

Adaptive Protein Evolution Detected by Neutrality Tests from Genomic Studies

Clark, et al. 2003 reports positively selected genes in the human lineage since its divergence from the chimpanzee lineage about six million years ago

Sabeti, et al. 2007 reports genes subject to comparatively recent position selection within humans

Bakewell, et al. 2007 reports that more genes underwent positive selection in chimpanzees than in humans since their separation, probably because of a larger effective population size and hence more effective selection in the former than the latter

Qiu, et al. 2012 reports positively selected genes related to hypoxia and hence highaltitude adaptation in the yak genome

Huang, et al. 2012 identifies fiftyfive selective sweeps during rice domestication

Raffaele, et al. 2010 reports positively selected genes in the fungal pathogen *Phytophthora infestans* that causes potato blight

Fraction of Adaptive AminoAcid Substitutions

Smith and EyreWalker 2002 proposes that one can estimate the fraction of aminoacid substitutions that are adaptive by

$$\alpha = 1 - (D_S P_N)/(D_N P_S)$$

EyreWalker 2006 reviews the estimates of α in a number of species, finding that it varies from nearly zero in humans and Arabidopsis to over 50 percent in *Drosophila* and some microbes and viruses

Using this method, Sawyer, et al. 2007 shows that approximately 95 percent of aminoacid substitutions in *Drosophila* is adaptive

The finding in *Drosophila* and several microbes that most amino-acid substitutions in a genome have been driven by positive selection seriously challenges the neutral theory

Other Types of Molecular Adaptation Detected by Neutrality Tests

In addition to the adaptive evolution of protein sequences, adaptive evolution of promoters that affect gene expression has also been reported. The best known examples include the human prodynorphin gene promoter reported in Rockman, et al. 2005 and human lactase gene promoter reported in Tishkoff, et al. 2007

The authors of Haygood, et al. 2007 have conducted a genomewide scan of human promoters to detect signals of positive selection

Gene loss may also be subject to positive selection, as reported in Wang, et al. 2006 for human *CASPASE12* and reported in MacArthur, et al. 2007 for human ACTN3

In addition to point mutations, insertions/deletions may be subject to positive selection, as reported in Podlaha and Zhang 2003 for mammalian CATSPER1, via the comparison between insertion/deletion substitution rates in coding and noncoding regions

Molecular Adaptations Revealed by Other Methods (detection of positive selection in molecular evolution)

Parallel and convergent aminoacid substitutions, whether they exceed the neutral expectations, e.g. parallel proteinsequence evolution hearing gene *Prestin* of echolocating bats and whales

Gene expression noise

Gene expression noise, defined by the variation in the mRNA or protein expression level of a gene among isogenic individuals in the same environment, has been shown to be ubiquitous

High level of gene expression noise may be advantageous for some genes under certain conditions and provides evidence that yeast plasmamembrane transporters are subject to positive selection for elevated expression noise

In principle, nothing rejects neutrality more convincingly than the demonstration that a substitution improves organismal fitness

Neutral Evolution of Genomic Architecture (origin of genomic architecture, including gene number, intron content, transposable element content, gene structure)

Evidence for the role of nearly neutral mutations in the origin of genomic complexity

Mutationalhazard hypothesis

Many complex elements of the genome, such as introns, are slightly deleterious. As the organism size increases in longterm evolution, the effective population size decreases, making it possible for slightly deleterious mutations to fix by chance

These elements may be subsequently tinkered by evolution to provide apparent utility, such as the existence of enhancers in introns, e.g. evolution by gene duplication. It was commonly believed that gene duplication leads to the origin of new gene functions

Duplicate genes are initially retained because of subfunctionalization (i.e., subdivision of ancestral functions to daughter genes) via degenerate mutations rather than neofunctionalization (i.e., acquisition of new functions) by advantageous mutations, i.e. gene number in a genome can increase via a pure neutral process

Genome size increases as the effective population size decreases, and that intron number and size, halflife of duplicate genes, and number of transposable elements all increase with genome size

However, bacterial genome size and effective population size are actually positively correlated rather than negatively correlated, because genome size tends to decrease under relaxed selection as a result of mutational deletion bias in bacteria

Using fitness data and a series of gene expression and protein function data that gene duplication frequently leads to adaptation

on the basis of expressional, functional, and fitness data that the duplication of a key player in the yeast galactose use pathway was adaptive

Beyond the single gene level, it has been argued that the expansions/contractions of certain gene families were subject to positive selection, e.g. genomic study of mammalian gene families; vertebrate vomeronasal receptor gene families

Neutrality concept

The neutral theory asserts that alternative alleles at variable protein loci are selectively neutral. This does not mean that the locus is unimportant, only that the alternative alleles found at this locus are selectively neutral

e.g. Glucose-phosphate isomerase is an esssential enzyme. It catalyzes the first step of glycolysis, the conversion of glucose-6-phosphate into fructose-6-phosphate

Natural populations of many, perhaps most, populations of plants and animals are polymorphic at this locus, i.e., they have two or more alleles with different amino acid sequences

The neutral theory asserts that the alternative alleles are selectively neutral.

Selectively neutrality

<u>Selectively neutrality</u> means that the alternative alleles have no effect on physiology or fitness and that the selection among different genotypes at this locus is sufficiently weak

Pattern of variation is determined by the interaction of mutation, drift, mating system, and migration

This is equivalent to saying that

$$N_e s < 1$$

where N_e is the effective population size and s is the selection coefficient on alleles at this locus

Experiments in *Colias* butterflies, and other organisms have shown that different electrophoretic variants of GPI have different enzymatic capabilities and different thermal stabilities. In some cases, these differences have been related to differences in individual performance

If populations of *Colias* are large and the differences in fitness associated with differences in genotype are large,

if $N_e s > 1$

then selection plays a predominant role in determining patterns of diversity at this locus, i.e., the neutral theory of molecular evolution would not apply

If populations of *Colias* are small or the differences in fitness associated with differences in genotype are small, or both, then drift plays a predominant role in determining patterns of diversity at this locus, i.e., the neutral theory of molecular evolution applies

The rate of molecular evolution

Rate of allelic substitution, under the hypothesis that mutations are selectively neutral

To get that rate two things are required:

the rate at which new mutations occur
and
the probability with which new mutations are fixed

rate of substitution = rate of mutation × probability of fixation

$$\lambda = \mu_0 p_0$$

Here, 'rate of mutation', is the number of new mutations that occur in any one generation

In a diploid population of size N, there are 2N gametes. The probability that any one of them mutates is just the mutation rate, μ, so,

$$\mu_0 = \mu$$

To calculate the probability of fixation, dynamics of alleles in populations is required.

Let's suppose that we're dealing with a single population. If the current frequency of an allele is p_0, what's the probability that is eventually fixed?

$$p_0$$

When a new mutation occurs there's only one copy of it, so the frequency of a newly arisen mutation is

$$p_0 = 1/\ 2N$$

Putting above two equations together, we find

$$\lambda = \mu_0 p_0$$
$$= (2N\mu)\ 1\ /2N$$
$$= \mu$$

If mutations are selectively neutral, the substitution rate is equal to the mutation rate

Since mutation rates are (mostly) governed by physical factors that remain relatively constant, mutation rates should remain constant, implying that substitution rates should remain constant

The prediction of a molecular clock follows directly from the hypothesis that mutations are selectively neutral

Diversity in populations (infinite alleles model)

Protein-coding genes consist of hundreds or thousands of nucleotides, each of which could mutate to one of three other nucleotides

It suggests that we could treat every mutation that occurs as if it were completely new, a mutation that has never been seen before and will never be seen again

This situation is well described by the infinite alleles model, One can calculate the equilibrium inbreeding coefficient for the infinite alleles model, i.e.,

$$f = 1/\ 4N_e\mu + 1$$

if the infinite alleles model is appropriate for molecular data, then f is the frequency of homozygotes we should see in populations and $1 - f$ is the frequency of heterozygotes So in large populations we should find more diversity than in small ones

The neutral theory is easily misinterpreted. It does NOT suggest:

That organisms are not adapted to their environments
That all morphological variation is neutral
That ALL genetic variation is neutral
That natural selection is unimportant in shaping **genomes**