

Molecular Population Genetics

I) Neutral Theory of Molecular Evolution

- Longstanding controversy in population genetics and evolutionary biology.
- Classical View – mutations are negative variants of the “wild-type.” Negative mutations are removed from the populations quickly. Positive mutations do occur but are fixed so quickly in the populations that the variation is never observed (i.e., previous “wild-type” and new positive mutation). Expectation is that little variation exists in natural populations.
- Neutralist View – after the modern synthesis a new school of neutral evolution was formed. Most mutations were neither positive nor negative but neutral in respect to fitness. Positive mutations occur and are fixed quickly. Negative mutations are purged. Later modified to neutral or nearly neutral (i.e., selection coefficients are very small).

OH 13.1

- To understand the difference between these two schools and why it was even a controversy, we need to understand how variation effects fitness.
- A) Genetic Load – when natural populations contain variation that is subjected to selection, the average population fitness may be less than its maximum value. The magnitude of the difference is referred to as the genetic load. Two types:
- 1) Mutational Genetic Load – reduction due to mutation alone.

$$L = \frac{W_{opt} - \bar{W}}{W_{opt}}$$

- Consider a single locus with 2 alleles (A and a) in a haploid population where $F(A) = p$ and $F(a) = q$ and fitnesses of 1 and $1 - s$ respectively.

- All “A” genotypes live (i.e., p of them)
- 1 – s of the “a” genotypes live [i.e., (1 – s)q of them]
- “s” of the “a” genotypes die (i.e., sq of them)

On a population level:

$$p + q(1 - s) \text{ live}$$

$$= p + q - qs = 1 - sq$$

$$\therefore \text{ in any generation: } \frac{\# \text{ dying}}{\# \text{ surviving}} = \frac{sq}{1 - sq}$$

Over many generations until the fixation of allele “A” the cost of this is:

$$C = \sum_{i=1}^{\text{fixation}} \frac{sq_i}{1 - sq_i}$$

In words, C = the amount surviving individuals must increase reproduction to maintain population size. The effect of this can be particularly large depending on population size. If populations are small then each per individual increase may be larger than biologically possible.

- 2) Segregational Load – if there is variation in natural populations it must be being maintained by balancing selection or overdominance.

Remember:

	AA	Aa	aa
Frequency	p^2	$2pq$	q^2
Fitness	$1 - s$	1	$1 - t$

AND:

$$\hat{p} = \frac{t}{s+t} \quad \text{and} \quad \hat{q} = \frac{s}{s+t}$$

Therefore:

$$\begin{aligned} F(\hat{H}) &= 2 \frac{t}{s+t} \frac{s}{s+t} \\ &= \frac{2st}{(s+t)^2} \end{aligned}$$

AND:

$$\begin{aligned} \hat{W} &= 1 - \frac{st}{s+t} \\ L &= \frac{st}{s+t} \end{aligned}$$

In words, the genetic load is the reduction in mean population fitness at equilibrium relative to the most fit population (i.e., for heterozygote advantage the most fit population would be 100% heterozygotes).

This is for one locus. For multiple loci the segregational load is:

$$\hat{W} = (1 - L)^i$$

over even a few loci evolving under heterozygote advantage the mean population fitness would be considerably lower than optimum (i.e., all individuals heterozygous at all loci). The amount that individuals can increase reproduction to compensate for this genetic load would likely be too small and the population would go extinct over time.

B) Neutral Theory of Molecular Evolution – strictly or nearly neutral view of genetic variation.

- 1) all alleles (or mutations) have the same probability of fixation, which is a function of their frequencies.**

new mutations:

$$F(\text{new mutant}) = \frac{1}{2N}$$

$$P(\text{fixation new mutant}) = \frac{1}{2N}$$

- 2) the number of new mutants in a population at any generation is a function of the population size and the mutation rate (# mutations per gene per generation)**

$$P(\text{new mutant}) = 2N\mu$$

- 3) Neutral Evolutionary Rate (k) – the rate at which new mutations are fixed in a population which is equal to the:**

P(new mutant) and P(fixation of new mutant)

$$= \frac{1}{2N} \times 2N\mu$$

$$= \mu$$

So the rate of neutral evolution is equal to the mutation rate and not population size (or anything else)! Time between mutations is, on average $\frac{1}{\mu}$. Population size only effects the time that it takes a mutation to go to fixation not the rate at which new mutations fix.

OH 13.2

II) Rates of Change

A) DNA sequence:

Probability of mutating a nucleotide is α

Since there are 4 different nucleotides the combined probability is 3α and the probability of not mutating is $(1 - 3\alpha)$.

$$P(A_{t+1}) = (1 - 3\alpha) P(A_t) + \alpha(1 - P(A_t))$$

or in words, the probability of still being A is equal to the probability that the nucleotide is A and did not mutate plus the probability that the nucleotide is not an A but mutates to an A.

The probability that two sequences separated for t generations are still the same:

$$P_{AA} = \frac{1}{4} + \frac{3}{4} e^{-8\alpha t}$$

and the probability that the sites are different is:

$$d = \frac{3}{4} (1 - e^{-8\alpha t})$$

This, however, is not a linear increase with time. As two sequences diverge by mutation there is an increasing probability that mutation will occur to a site that has already been mutated. After enough time, it is increasingly likely that new mutation will increase similarity between sequences by multiple substitution (substitution saturation) at the same site.

OH 13.3

B) Nucleotides:

OH 13.4A, 13.4B

C) Codons

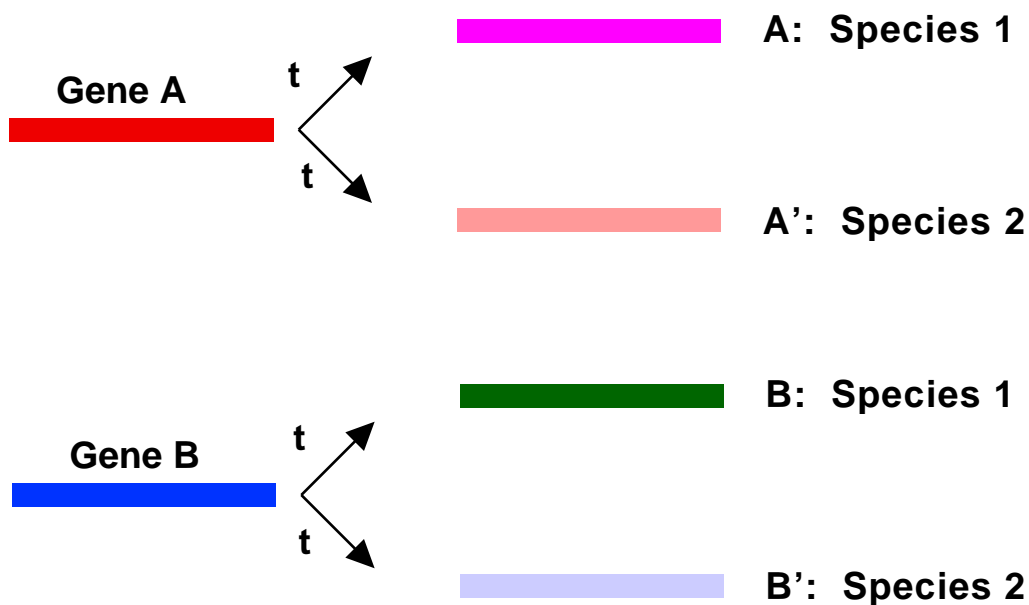
OH 13.5

- 1) Synonymous versus Nonsynonymous substitutions.
- 2) Codon Bias

OH 13.6

III) Test of Neutrality

A) HKA – Hudson, Kreitman, and Aguade test.



As we have seen, assuming neutrality:

$$\hat{H} = \frac{4N\mu}{4N\mu + 1} = \frac{\theta}{\theta + 1}$$

For gene A then we can estimate θ_A (heterozygosity of the population) and assume a neutral mutation of μ_A

\therefore the differences between A and A' = $K_A = \mu_A 2t$

and then $\mu_A = \frac{K_A}{2t}$

From these:

$$\theta_A = 4N \frac{K_A}{2t}$$

And:

$$t = 4N \frac{K_A}{2\theta_A} = \text{time since divergence of the two species}$$

turning to gene B:

We now have an estimate of the observed and expected (based on the time estimate in gene A) level of sequence variation for gene B which can be tested with a χ^2 .

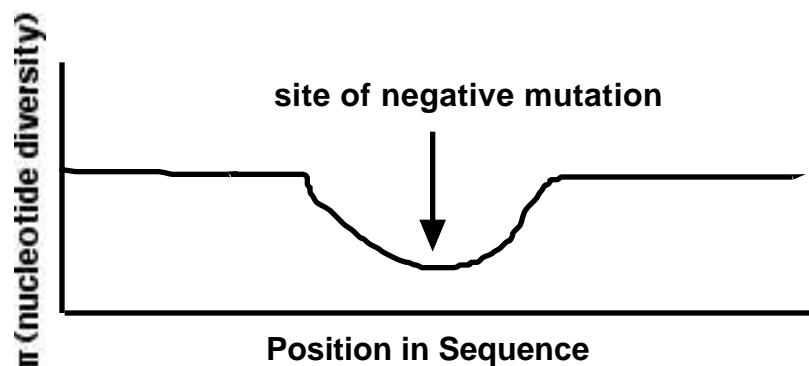
e.g. HKA found that there were too many neutral mutations in ADH of *D. melanogaster*.

OH 13.7

IV) Effects of Recombination on Sequence Variation and Mutations

A) Background Selection :

- 1) Generally mutation is negative and reduces the number of offspring from the mutated individual. This is similar to a lower effective population size (i.e., reduced genetic variation).
- 2) Recombination – recombination between mutated and non-mutated sites releases the flanking region from the effects of the lower effective population size.



- B) Selective Sweeps** – when positive mutations occur they generally are fixed in the population quickly. When they are, the flanking sequences also are fixed in the population even when neutral. With recombination, the flanking sites are, again, released, but mutations occurring to linked region will accumulate. Again, however, this is a reduction in the effective population size.
- C) Balancing Selection** – heterozygote advantage allows multiple alleles to be maintained in the population longer than they would under a neutral mutation. The amount of variation at linked sites is generally higher because of the increased longevity of these regions relative to unlinked flanking regions.

OH 22.5

V) Detecting Recombination

A) Issues

- 1) Coadapted gene complexes** – on a genomic basis, recombination is important because it creates the opportunity for and breaks down coadapted gene complexes.
- 2) Anastomosis** – when viewing the relationship among a group of individuals, linked (i.e., no recombination) characters should all produce the same network of relationships. Unlinked site, however, may not.

OH 13.8

VI) Multi-gene families – single copy DNA is probably nonexistent due to duplication of genes into related groups of sequences (i.e., families).

A) Duplication and divergence before speciation

OH 13.9A, 13.9B

B) Gene Conversion

OH 13.10A, 13.10B, 13.10C

C) Biased Gene Conversion and Concerted Evolution

OH 13.10D

D) Stochastic Lineage Sorting

OH 13.11, 13.12, 13.13