• A selected allele which increases its bearer’s relative fitness by \( s \) has just been introduced by mutation
• Strongly advantageous alleles \( s >> 1/(2N_e) \) have a probability \( s \) of fixing in population
• Very deleterious alleles \( |s| >> 1/(2N_e) \) can never fix
• Alleles with \( |s| < 1/(2N_e) \) are effectively neutral & have \( 1/(2N_e) \) fixing in population.
In large populations selection can act on very small fitness benefits

Codon-usage bias

Frequencies of the 6 leucine codons in coding regions

Frequency of tRNAs in cells

Codon bias correlates
With expression level in *Drosophila* *mel*.

This effect is not strongly seen in vertebrates
The Molecular Clock

Observation: Rate of amino acid substitution in many, but not all, proteins surprisingly constant over time, i.e. evolve in a clock-like manner.

• Neutral theory Claim: Due to most amino acid replacement substitutions between species being neutral
Generation time effect

Strong generation time effect seen for least constrained sites. i.e. the changes most likely to be neutral, do not show a molecular clock in years. Non-syn. substitutions show much less of a generation time effect.
Generation time effect

Strong generation time effect seen for least constrained sites. i.e. the changes most likely to be neutral, do not show a molecular clock in years.
Non-syn. substitutions show much less of a generation time effect.

Neutral theory of molecular evolution cannot explain molecular clock measured in years. It predicts a molecular clock measured in generations, which is exactly what we see for changes to synonymous/non-coding sites but not what we see for changes to the protein.

So what explains the protein molecular clock?

Phylogram from Dog genome paper (Nature 2005)
Using mostly non-coding sequence data
Nearly Neutral theory proposed by Ohta
Perhaps much of molecular evolution
due to substitutions with selection coefficients $|s| \sim 1/(2N_e)$
Much to commend this idea, including fact that it may explain the protein molecular clock.

Effective Population size potentially determines levels of constraint

Hominids show weaker conservation of CNG (Conserved non-genic) than rodents.

Keightley et al 2005

Effective Population size potentially determines levels of constraint
Current status of views on molecular evolution

Most* of the substitutions/polymorphism in non-coding DNA & synonymous changes in organisms with large genomes are likely neutral, as are a reasonable fraction of non-synonymous changes.

Weakly deleterious alleles likely make a significant contribution to substitution rates in species (such as our own) with small effective population sizes.

There is increasing evidence that a reasonable fraction (>10%) of non-synonymous changes are driven by selection in species with large effective population sizes. But the selection coefficients may be very small $\sim 1/(2N_e)$. We have no clue what most of these do!

* --but not all
The Strength of Selection against Neanderthal Introgression

Ivan Juric, Simon Aeschbacher, Graham Coop

Published: November 8, 2016 • http://dx.doi.org/10.1371/journal.pgen.1006340

http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006340

See also Harris & Nielsen

Why You Don’t Have Much Neanderthal DNA in Your Genome


https://www.flickr.com/photos/hmnh/3033749380/
~7000 Exonic sites harbouring deleterious alleles

\[ s_{EUR} = 4.1 \times 10^{-4} \]

\[ s_{ASN} = 3.5 \times 10^{-4} \]
Genomic Regions with: Less genes → More genes

Exonic base pairs per cM
Neanderthal and Human divergence

600,000 yrs

$N_e = 2k$

$N_e = 10k$

Fraction of sites vs. log10(s)

Neanderthal fixed sites

Human fixed sites

$1/2N_e$
Implications

• Levels of introgression are shaped by weakly deleterious alleles.
  – Likely due to nearly neutral alleles drifting up in N.

• Despite being weakly deleterious in aggregate fitness drop many be profound.

• Differences in population size common across closely related species, so these dynamics may be generally important
Diversity-reducing selection

Best studied models:

• hitch-hiking models (Maynard-Smith and Haigh 1974; Kaplan et al. 1989)

• background selection models (Charlesworth et al. 1993; Hudson and Kaplan 1994).
The hitchhiking effect: Indirect impact of selection on patterns of linked diversity

New beneficial mutation Arises on a particular genetic background (haplotype).

If there is no recombination
Indirect impact of selection on patterns of diversity

The hitchhiking effect
When recombination occurs

New beneficial mutation
Arises on a particular genetic background (haplotype).

A haplotype a set of alleles physically linked on a chromosome
Indirect impact of selection on patterns of diversity

A full sweep:
The selected allele has reached fixation.

- Reduces levels of polymorphism (i.e. heterozygosity)
- Stronger selection/lower recombination leads to reduction over larger genomic region.
An example of the hitchhiking effect in humans

A selected allele has swept to high frequency in Eurasia. Reducing levels of heterozygosity in a broad genomic region.
At the selected locus, two subpopulations:

- **Derived allele**
- **Ancestral allele**

The diagram shows the frequency of the derived allele over time, with the frequency ranging from 0 to 1 along the x-axis and time on the y-axis. The graph indicates a decrease in frequency over time.
A sample from the neutral locus fully linked to selected locus

Frequency of derived allele

Hudson et al '88
Fig. 6 A multilocus scan of microsatellite variation in a southeast Asian population of the malaria parasite *Plasmodium falciparum* reveals evidence for a selective sweep on chromosome 4. Microsatellite variation was markedly reduced within an interval of c. 100 kb centered on *dhfr*, a gene involved in drug resistance [reproduced from Nair *et al.* (2003)].
Genomic patterns

Low recombination rate region

High recombination rate region

neutral polymorphism level

chromosomal position

~100 kb  ~100 kb

~10 kb  ~10 kb
Polymorphism levels are positively correlated with recombination, likely due to hitchhiking (like) effect.

Genome-wide evidence
For directional selection
Genomic effects of directional selection

Heterozygosity levels are positively correlated with recombination across the genome.

In Drosophila

Begun and Aquadro 1992, Nature

Heterozygosity

Rate of Recombination

Strong genome-wide evidence for hitchhiking effect and
A wide role for directional selection in reducing levels of polymorphism

Same effect seen in human polymorphism data

Cai et al 2009
The effect of a selective sweep on completely linked neutral variants

Favored alleles fix quickly: $\sim 2\log(2N_e)/s$ compared to $4N$ generations for a neutral allele.
E.g., for $N_e = 10000$, $s = 0.01$, that is 1,980 generations instead of 40,000 generations.