Chapter 8 – Molecular Evolution

Neutral/Nearly Neutral Theory

Measuring Divergence & Polymorphism

The Molecular "Clock"

Variation in Molecular Rates

Tests for Deviation from Neutral Expectations

Molecular Evolution at Linked Loci/Sites

Molecular Population Genetics

- increasing complexity of genetic data
  - "Aa" to allozymes to RFLPs to DNA sequences
- the same evolutionary processes (mutation, drift, selection, gene flow) considered in the "Modern Synthesis" are responsible for the patterns of variation observed in DNA sequences
- theory developed for two alleles at a locus is generally applicable to DNA sequence data, although the additional information contained in sequences provides added information…

The Neutral Theory

- Kimura (1968) - ***most*** polymorphism at the molecular level is selectively neutral
  - ideas developed during the allozyme era, but even more relevant for DNA sequences
- strongly deleterious or advantageous mutations are expected to be eliminated or fixed quickly and therefore not contribute much to segregating polymorphism
  - mutation rate μ is effectively the *rate of neutral mutation*

Neutral Expectations…

- …with constant population size and mutation

nucleotide diversity

$$E(\Pi) = \theta = 4N\mu$$

 # segregating sites

$$E(S) = \theta\left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots + \frac{1}{k-1}\right)$$

homozygosity

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

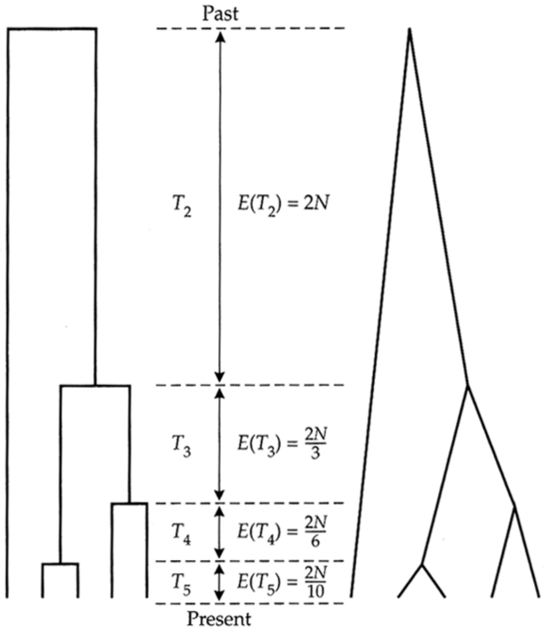
 # unique alleles

$$E(k) = 1 + \frac{\theta}{\theta+1} + \frac{\theta}{\theta+2} + \ldots + \frac{\theta}{\theta+n-1}$$

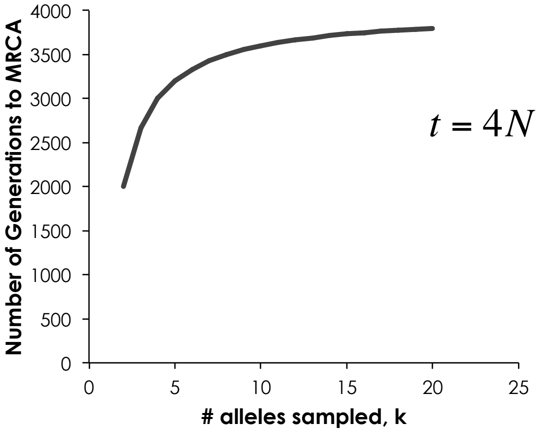
allele frequency distribution

$$\Pr\{a_1,\ldots,a_n\} = \frac{n!}{\theta(\theta+1)\ \ldots\ (\theta+n-1)} \prod_{j=1}^{n} \frac{\theta^{a_j}}{j^{a_j} a_j!}$$

FIGURE 3.15 Two completely equivalent ways of illustrating the coalescences in a gene tree. On the left, the coalescent events are represented as horizontal lines, on the left they are represented as nodes. In any each generation, if there are k alleles present, the expected time back to the next coalescence is given by $4N/[k(1-k)]$. For example, starting with five alleles, the expected time back to the first coalescence is $4N/[(5)(4)] = 2N/10$. Note that the successive times get longer. When there are only two alleles, the time back to the final coalescence is $2N$ generations.



Time to Coalescence



$$t = 4N\left(1-\frac{1}{k}\right)$$

MRCA = most recent common ancestor, N = 1,000

Basic Neutral Theory Principles

- at mutation/drift equlibrium (gain and loss of alleles due to mutation and drift are equal), expected homozygosity is: $$\frac{1}{(4N_e\mu+1)}=\frac{1}{\theta+1}$$
- and expected heterozygosity is: $$1-\frac{1}{\theta+1}=\frac{\theta}{\theta+1}$$
- so larger populations should have greater genetic diversity

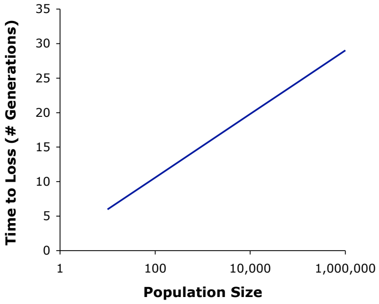
Mutation vs. Fixation/Substitution

Fixation/Substitution

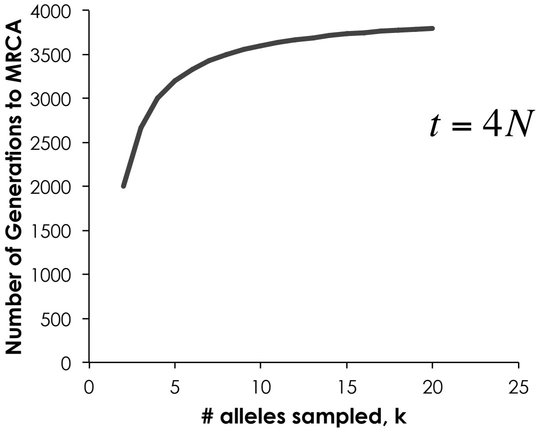
- average time to fixation (for alleles that eventually become fixed) is $4N_e$ generations
- average time to loss of new mutations is

$$\frac{2N_e}{N}\ln(2N) = 2\ln(2N)$$

generations



Time to Coalescence



$$t = 4N\left(1 - \frac{1}{k}\right)$$

MRCA = most recent common ancestor, N = 1,000

Probability of Fixation & Rate of Evolution

- new mutations have a frequency of $1/(2N)$ and therefore have a probability of fixation of $1/(2N)$
- new mutations become fixed at a rate equal to μ (see next slide)
  - mutation rate equals "substitution" rate
  - neutral rate is independent of population size!
- average time interval between fixation of new mutations is $1/\mu$

Neutral Expectations…

- the rate of neutral evolution is independent of population size
  - substitution rate (k) equals mutation rate

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

Average time to fixation/substitution

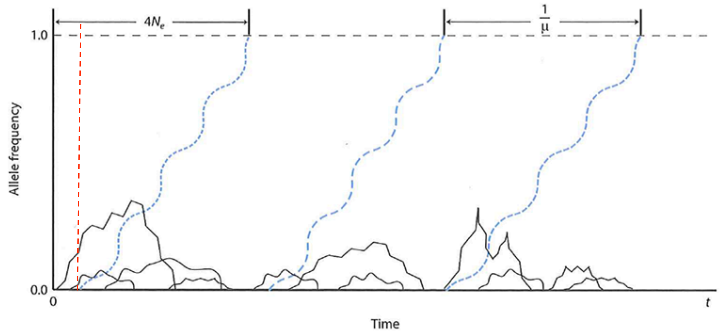
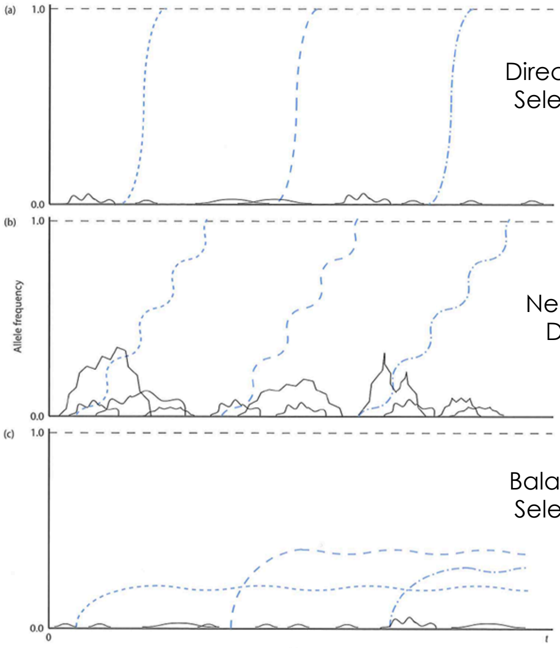


Figure 8.2 The fate of selectively neutral mutations in a population. New mutations enter the population at rate μ and an initial frequency of $\frac{1}{2N}$. Allele frequency is a random walk determined by genetic drift. The time that a new mutation segregates in the population, or the dwell time of a mutation, depends on the effective population size. However, the chance that a new mutation goes to fixation (equal to its initial frequency) is also directly related to the effective population size. These two effects of the effective population size cancel each other out for neutral alleles. The neutral theory then predicts that the rate of fixation is μ and therefore the expected time between fixations is $1/\mu$ generations. For that subset of mutations that eventually fix, the expected time from introduction to fixation is $4N_e$ generations. After Figure 3.1 in Kimura (1983a).

Fate of New Mutations



Directional Selection

Neutral Drift

Balancing Selection

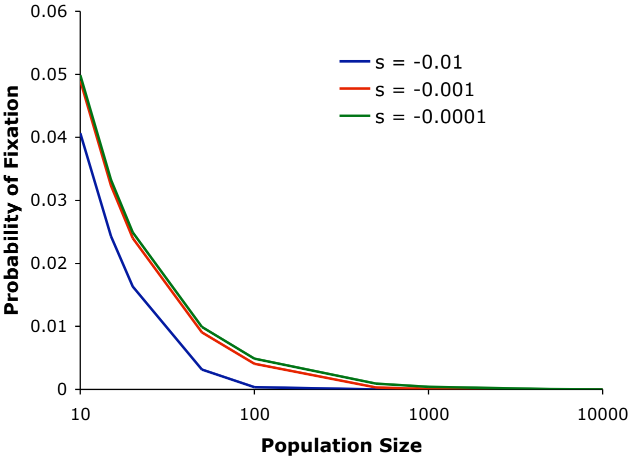
Figure 8.3 The dwell time for new mutations is different if fixation and loss is due to genetic drift or natural selection.

Nearly Neutral Theory

- what happens in small populations when selection is weak?
  - changes in allele frequency due to drift and selection are approximately equal $|2Ns| \approx 1$
- probability of fixation for a new, “nearly neutral” allele:

$$\Pr(A \text{ fixed}) = \frac{2s}{1-e^{-4Ns}}$$

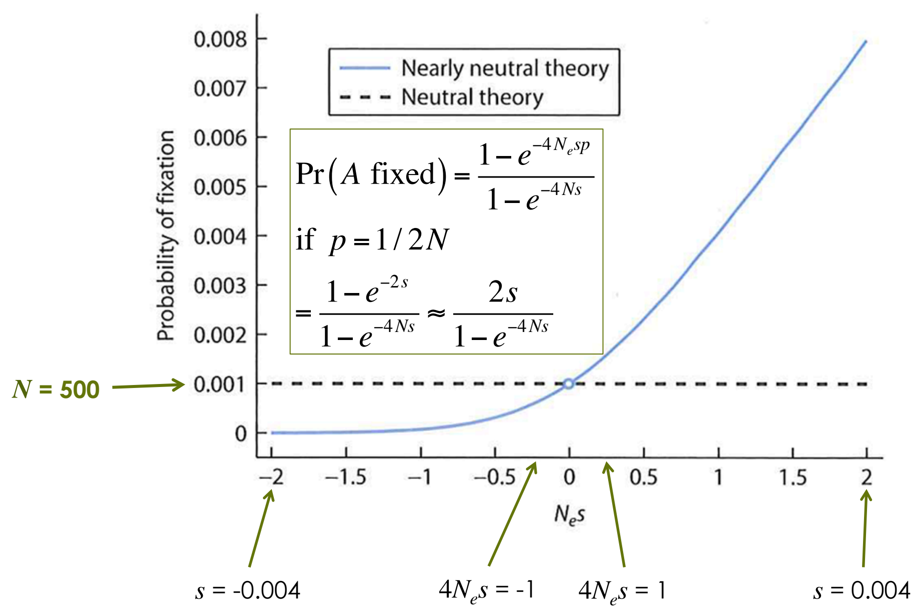
$$w_{AA} = 1 + \underline{s}, \quad w_{Aa} = 1 + \underline{s/2}, \quad w_{aa} = 1$$



$$\Pr(A \text{ fixed}) = \frac{2s}{1-e^{-4Ns}}$$

What qualifies as nearly neutral?

- Hamilton: $2s = 1/2N_e$ or $4N_es = 1$
  - value at which "the processes of genetic drift and selection are **equal**"
- Hartl & Clark: $|2Ns| \approx 1$
- Hedrick: $s < 1/(2N)$ <u>or</u> $2Ns < 1$
- Ohta & Gillespie (1996): $s \approx 1/N$ <u>or</u> $Ns \approx 1$



$$\Pr(A\ \text{fixed}) = \frac{1-e^{-4N_esp}}{1-e^{-4Ns}}$$

if $p = 1/2N$

$$= \frac{1-e^{-2s}}{1-e^{-4Ns}} \approx \frac{2s}{1-e^{-4Ns}}$$

N = 500

s = -0.004

$4N_es$ = -1

$4N_es$ = 1

s = 0.004

Nearly Neutral Theory - Summary

- the rate of neutral evolution is independent of population size
  - substitution rate equals mutation rate

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

- in contrast, the fate of nearly neutral mutations depends on population size
  - when N is small, the effect of genetic drift can be comparable to that of selection, making slightly deleterious mutations "effectively neutral"

$$|2Ns| \approx 1$$

- thus, lineages experiencing small population size should accumulate both neutral and nearly neutral mutations, leading to a faster rate of sequence evolution

---

Woolfit & Bromham 2003 Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. *MBE* 20:1545-1555.

- higher rate in endosymbiotic bacteria interpreted as a consequence of nearly neutral evolution