

Chapter 8 – Molecular Evolution

Neutral/Nearly Neutral Theory
Measuring Divergence & Polymorphism
Sequence Divergence & The Molecular “Clock”
Variation in Molecular Rates
Tests for Deviation from Neutral Expectations
Molecular Evolution at Linked Loci/Sites

Sequence Divergence

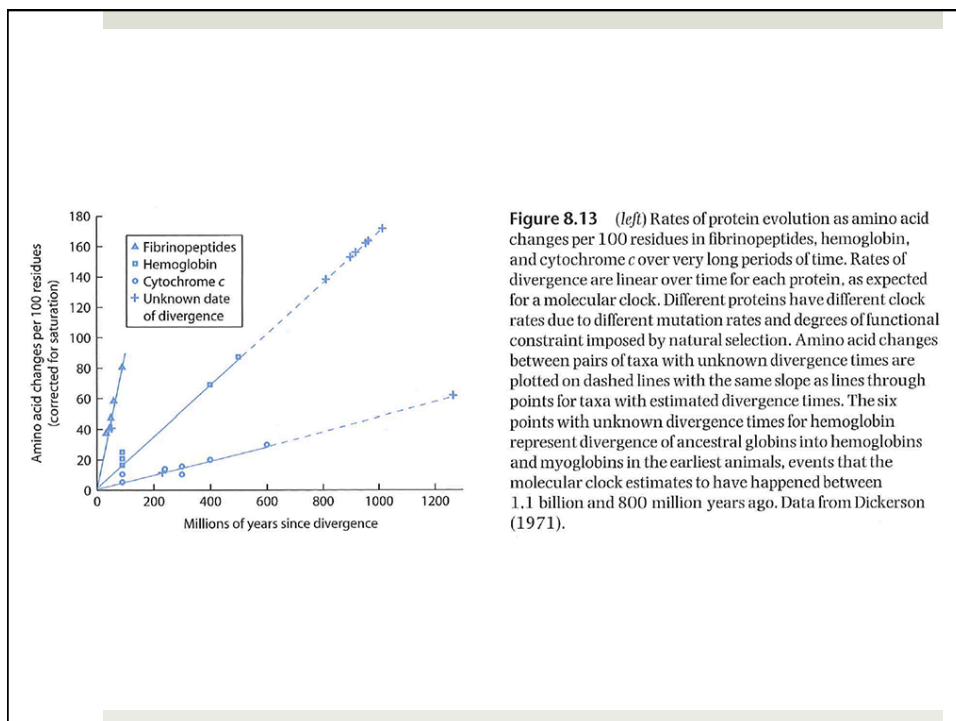
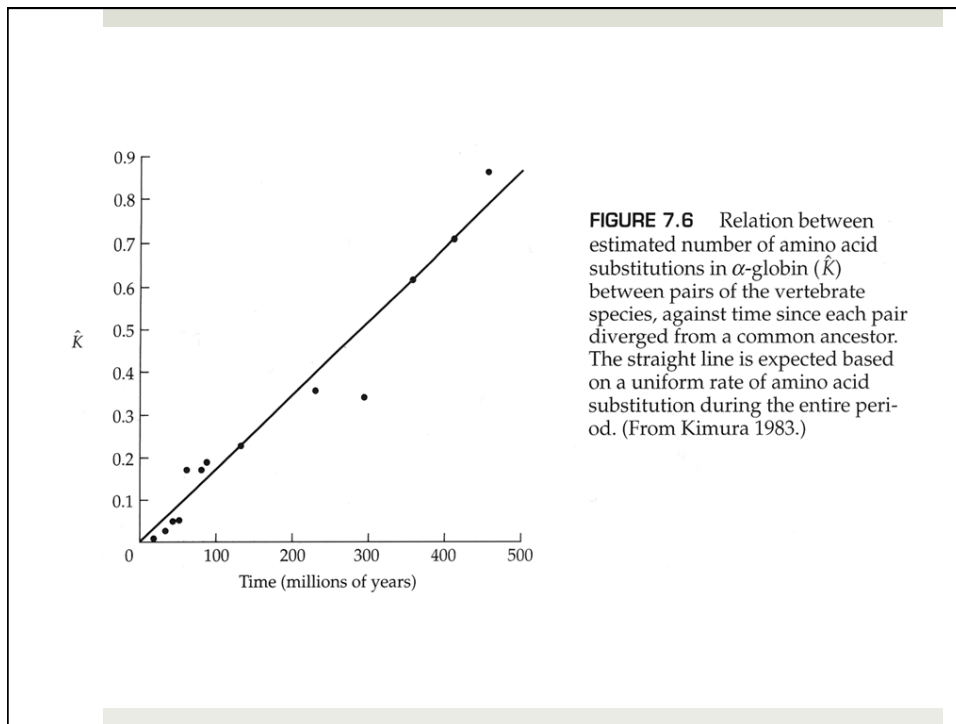
- ❖ simple genetic distance, d = the proportion of sites that differ between two aligned, homologous sequences
- ❖ given a constant mutation/substitution rate, d should provide a measure of time since divergence
 - ❖ but this is greatly complicated by **multiple hits** (homoplasy)
 - ❖ corrected distance measures account for the fact that there are **not** an infinite number of sites in a sequence

Expected sequence divergence

- ❖ for neutral polymorphisms, substitution rate = mutation rate
- ❖ thus, for two diverging lineages... $k = 2T\mu$
 - ❖ where k = the number of substitutions observed between two species and T is the time since divergence
 - ❖ note that T can be measured either in years or generations
- ❖ solving for T ... $T = \frac{k}{2\mu}$
 - ❖ note that 2μ is often expressed as the "rate of sequence divergence" (i.e., twice the per lineage rate)

Rates and Dates: Divergence Time Estimates

- ❖ requires calibration with fossil or geological events
- ❖ typically assumes a "molecular clock"
 - ❖ Zuckerkandl & Pauling (1962)
- ❖ but new methods allow a relaxation of the molecular clock assumption



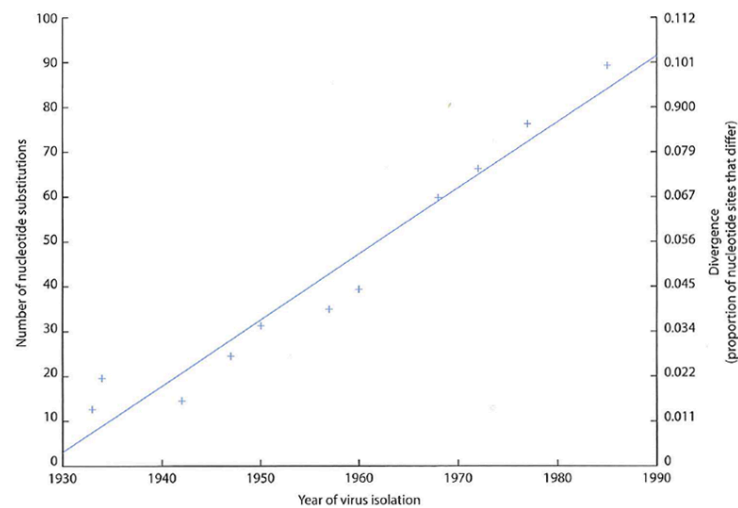
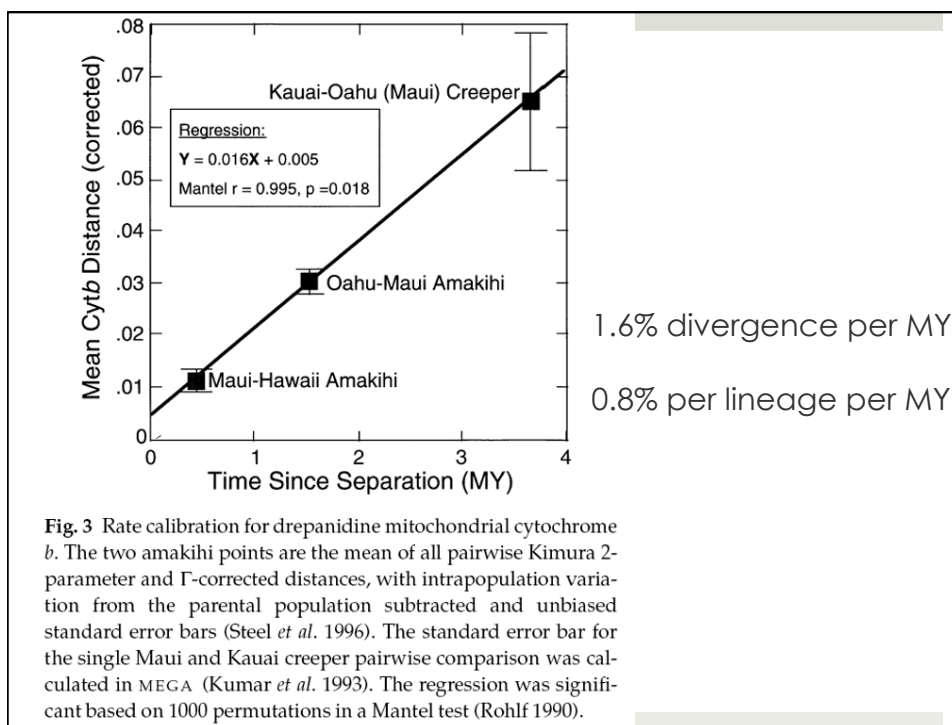
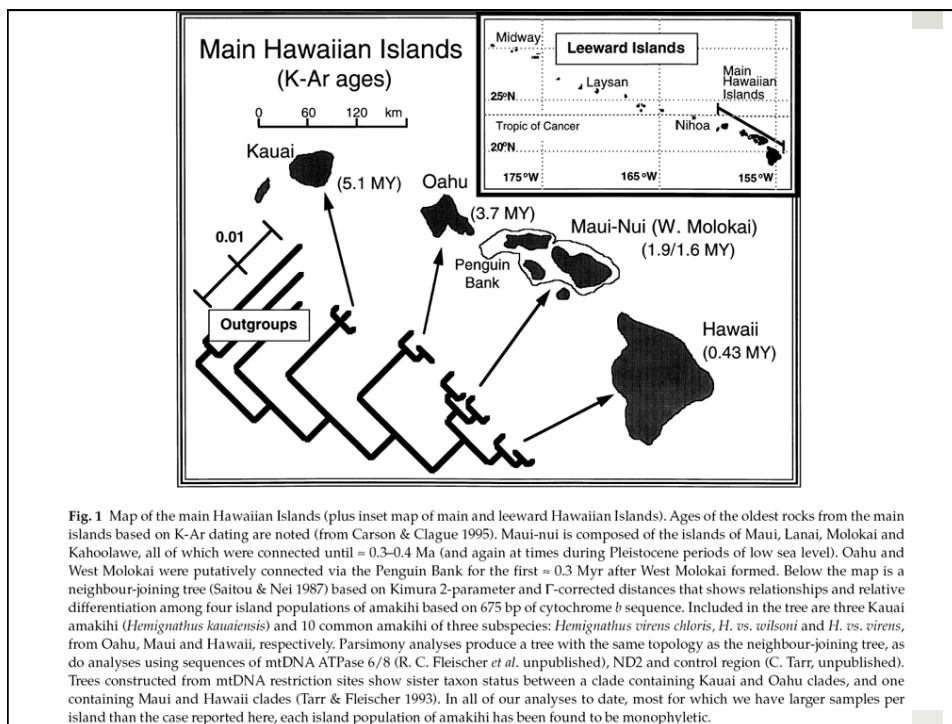


Figure 8.12 Rates of nucleotide change in the NS gene that codes for “nonstructural” proteins based on 11 human influenza A virus samples isolated between 1933 and 1985. The number of years since isolation and DNA sequence divergence from an inferred common ancestor are positively correlated. The pattern of increasing substitutions as time since divergence increases is expected under the molecular clock hypothesis. The observed rate of substitution was approximately 1.9×10^{-3} substitutions per nucleotide site per year, a very high rate compared to most genes in eukaryotes. The line is a least-squares fit. Data from Buonagurio et al. (1986).

Fleischer *et al.* 1998. Evolution on a volcanic conveyor belt: using phylogeographic reconstructions and K-Ar based ages of the Hawaiian Islands to estimate molecular evolutionary rates. *Mol. Ecol.* 7:533-545.





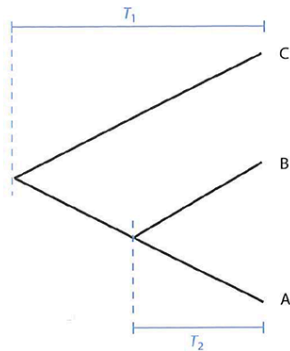


Figure 8.14 A schematic phylogenetic tree that can be used to date divergence events under the assumption of a constant rate of divergence over time or a molecular clock. T_1 is the time in the past when species C and the ancestor of species A and B diverged. T_2 is the time in the past when species A and B diverged. If either T_1 or T_2 are known, the rate of molecular evolution per unit of time can be estimated from observed sequence divergences. This rate of divergence can then be used to estimate the unknown amount of time that elapsed during other divergences.

suppose T_1 is known...

$$\mu = \frac{1}{2} \left(\frac{K_{AC}}{2T_1} + \frac{K_{BC}}{2T_1} \right)$$

Problems with dating...

- ❖ uncertainty in calibration points
- ❖ fossil evidence provides lower bound on age only
- ❖ variance of genetic distance estimates
- ❖ “saturation” of genetic distances
- ❖ extrapolation outside of calibrated range
- ❖ ancestral polymorphism
- ❖ **variation in substitution rate among lineages**