

Molecular clocks

Comparative Genomic Analyses

Rui Borges
Vetmeduni Vienna

The molecular clock

The genetic distance of two sequences coding for the same protein, but isolated from different species, seems to increase linearly with the divergence time.

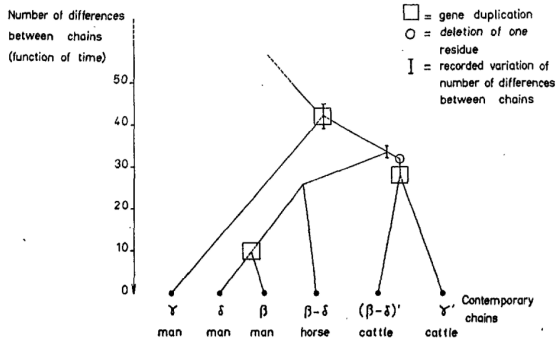


FIG. 4. Probable evolutionary relationship of some mammalian hemoglobin chains.

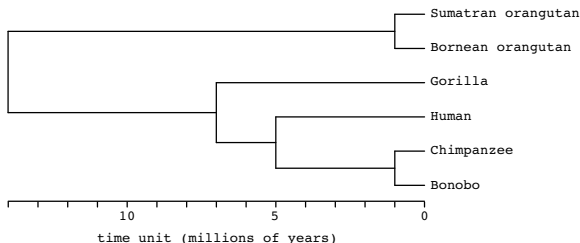
Zuckerkandl & Pauling (1965)

The molecular clock

Exercise

We want to know if a certain gene shows a constant rate of evolution among six great apes.

	1	11	21	31	41
Bonobo	ACTGGTATAG	GTGGCACGCA	GCCCGTTATA	AGACGTTTCA	TAGCAAAGGG
Chimpanzee	ACTGGTATAG	GTGGCACGCA	GCCCGTTATT	AGACGTTTCA	TAGCAAAGGG
Human	ACTGGTATCC	GTAGCACGCA	GCCCGTTATT	AGACGTTTCA	TCGCAAAGGG
Gorilla	ACTGGTATCC	GTAGCAAGCA	GCCCGCTTTG	AGAGCTTTCT	TCGCAAAGGG
Bornean orang	ATTGGCATCC	CTAGCAAGCC	GCCCGGTTTA	AGTCATTTC	TCGCAAAGGA
Sumatran orang	AGTGGTATCC	CTAGCAAGCC	GCCCGGTTTT	AGTCATTTC	TCGCAAAGGA

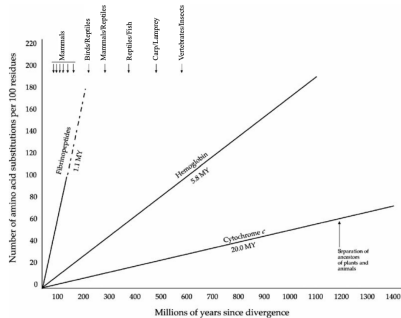


Plot the p-distance as a function of the divergence time for each pair of species and conclude on the existence of a molecular clock for this gene.

The molecular clock

The rate of evolution for any given protein is constant over time.

- ▶ implies the existence of a **molecular clock** ticking faster or slower for different genes
- ▶ more or less constant rate for any given gene among different phylogenetic lineages



Hartl & Clark (1997)

The molecular clock

The molecular clock is central to our understanding of evolutionary processes.

- ▶ a tool for estimating historical dates and rates of evolution
- ▶ provides a formal description of the substitution process
- ▶ allows for more accurate phylogenetic reconstructions
- ▶ deviations from clock-like behavior may reveal adaptive evolution, relaxed functional constraints, or changes in effective population size.

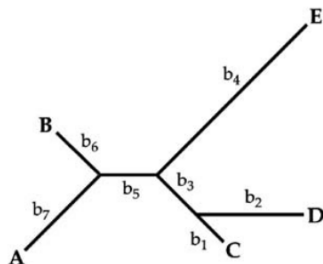
Testing the clock hypothesis

According to the molecular clock hypothesis

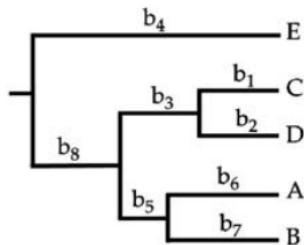
- ▶ two taxa sharing a common ancestor t years ago should have accumulated more or less the same number of substitutions during time t

Testing the clock hypothesis

Non-clock-like tree



Clock-like tree



- ▶ unrooted tree
- ▶ $2n - 3$ independent branches
- ▶ $b_1, b_2, b_3, b_4, b_5, b_6$ and b_7 need to be estimated

- ▶ rooted tree
- ▶ $n - 1$ independent branches
- ▶ b_1, b_3, b_4 and b_6 need to be estimated

Testing the clock hypothesis

The branch lengths of a phylogeny can be estimated under the clock or non-clock scenarios.

- ▶ the molecular clock is the null hypothesis and represents a special case of the more general alternative hypothesis assuming a different rate for each branch
- ▶ the LRT can be used to evaluate whether or not the sequences have been evolving at the same rate

Felsenstein (1988)

Testing the clock hypothesis

- ▶ Hypothesis

H_0 : global clock versus H_1 : non-clock

- ▶ Test statistic

$$LRT = 2(\ell_1 - \ell_0)$$

- ▶ Decision rule

LRT follows a χ^2 distribution with $n - 2$ degrees of freedom

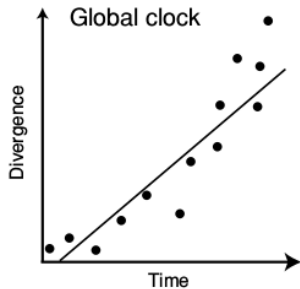
Relaxing the molecular clock

Empirical studies that employed rigorous statistical testing have revealed non-clock behavior in many gene sequences

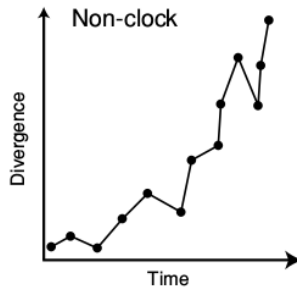
- ▶ potential sources of rate variation are generation times, replication and repair mechanisms and differences in natural selection

Bromham and Penny (2003)

Relaxing the molecular clock



- ▶ same rate for all the branches
- ▶ oversimplified and badly fitting model



- ▶ too many parameters
- ▶ the contribution of evolutionary time and rate is unidentifiable

Relaxing the molecular clock

The clock hypothesis could be relaxed by allowing separate rate parameters to different branches or collections of branches

- ▶ Local molecular clock model
Yoder & Yang (2000)
- ▶ comparison of local clock models with the global clock model possible via LRT
- ▶ specify the null hypothesis beforehand may be difficult

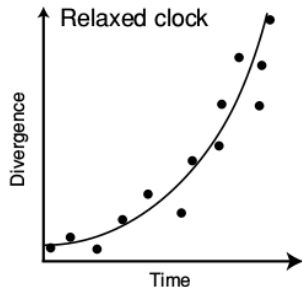


Relaxing the molecular clock

Bayesian treatment of clock models assume rates of evolution vary through time

(Kishino et al., 2001; Thorne et al., 1998)

- ▶ **autocorrelated relaxed clock**
rates across branches are autocorrelated
- ▶ **uncorrelated relaxed clock**
branch-specific rates are drawn independently from an underlying rate distribution



Applications of the molecular clock

- ▶ pathogen epidemiology
among others Korber et al. (2000) and Lemey et al. (2006)
- ▶ the study of the origin of the main types of animals
for review see Bromham & Penny (2003)
- ▶ studies of historical changes in mammalian population sizes
Shapiro et al. (2004)
- ▶ evidence against deliberate virus transmission in an HIV outbreak
de Oliveira et al. (2006)

Literature

The phylogenetic handbook by Lemey, Salemi and Vandamme (2009)
Cambridge University Press

- ▶ Chapter 11: sections 11.1, 11.2, 11.3, 11.5 and 11.6