**Course: Advanced Bioinformatics**

**Module title: Molecular Clock**

**Module no. : 51**

**Molecular clocks**

The molecular clock (based on the molecular clock hypothesis (MCH)) is a technique in molecular evolution that uses fossil constraints and rates of molecular change to deduce the time in geologic history when two species or other taxa diverged. It is used to estimate the time of occurrence of events called speciation or radiation. The molecular data used for such calculations are usually nucleotide sequences for DNA or amino acid sequences for proteins. It is sometimes called a gene clock or evolutionary clock.

**Non-constant rate of molecular clock**

Sometimes only a single divergence date can be estimated from fossils, with all other dates inferred from that. Other sets of species have abundant fossils available, allowing the MCH of constant divergence rates to be tested. DNA sequences experiencing low levels of negative selection showed divergence rates of 0.7–0.8% per Myr in bacteria, mammals, invertebrates, and plants. In the same study, genomic regions experiencing very high negative or purifying selection (encoding rRNA) were considerably slower (1% per 50 Myr).

In addition to such variation in rate with genomic position, since the early 1990s, variation among taxa has proven fertile ground for research too, even over comparatively short periods of evolutionary time (for example mockingbirds). Tube-nosed seabirds have molecular clocks that on average run at half speed of many other birds, possibly due to long generation times, and many turtles have a molecular clock running at one-eighth the speed it does in small mammals or even slower. Effects of small population size are also likely to confound molecular clock analyses. Researchers such as Francisco Ayala have more fundamentally challenged the molecular clock hypothesis. According to Ayala's 1999 study, five factors combine to limit the application of molecular clock models:

* Changing generation times (If the rate of new mutations depends at least partly on the number of generations rather than the number of years)
* Population size (Genetic drift is stronger in small populations, and so more mutations are effectively neutral)
* Species-specific differences (due to differing metabolism, ecology, evolutionary history.)
* Change in function of the protein studied (can be avoided in closely related species by utilizing non-coding DNA sequences or emphasizing silent mutations)
* Changes in the intensity of natural selection.

Molecular clock users have developed workaround solutions using a number of statistical approaches including maximum likelihood techniques and later Bayesian modeling. In particular, models that take into account rate variation across lineages have been proposed in order to obtain better estimates of divergence times. These models are called relaxed molecular clocks[15] because they represent an intermediate position between the 'strict' molecular clock hypothesis and Joseph Felsenstein's many-rates model[citation needed] and are made possible through MCMC techniques that explore a weighted range of tree topologies and simultaneously estimate parameters of the chosen substitution model. It must be remembered that divergence dates inferred using a molecular clock are based on statistical inference and not on direct evidence.

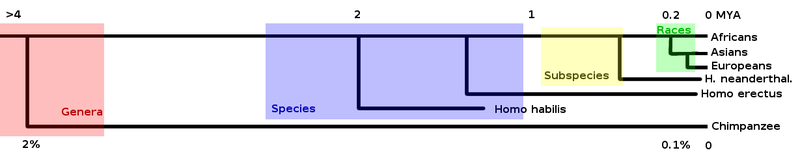
The molecular clock runs into particular challenges at very short and very long timescales. At long timescales, the problem is saturation. When enough time has passed, many sites have undergone more than one change, but it is impossible to detect more than one. This means that the observed number of changes is no longer linear with time, but instead flattens out.

At very short time scales, many differences between samples do not represent fixation of different sequences in the different populations. Instead, they represent alternative alleles that were both present as part of a polymorphism in the common ancestor. The inclusion of differences that have not yet become fixed leads to a potentially dramatic inflation of the apparent rate of the molecular clock at very short timescales [16][17].

Evolutionary biologists have revolutionized their field using a combination of DNA sequencing and bioinformatics. By comparing inherited DNA substitution mutations, or single nucleotide polymorphisms (SNPs), it is possible to establish relationships between different population groups within the same species or between different closely related species. When two new species develop from a common ancestor each will inherit and continue to accumulate a unique set of random mutations. Assuming the mutations accumulate at a constant rate, the number of mutations will be proportional to the length of time that two groups have been separated. In other words the number of mutations will be equivalent to a set period of time – a molecular clock. Knowing the rate at which the changes occur and the number of differences between two organisms allows you to estimate when the separation from the common ancestor occurred.

Before using the clock, it has to be calibrated. For example, by extrapolating from fossil evidence and DNA sequences from other primates it is thought that chimpanzees and humans diverged from a last common ancestor about 5 to 6 million years ago. By measuring the number of sequence differences between chimpanzees and humans it is possible to set the clock. However, this is a controversial process, for example it relies on having fossil evidence which can be accurately dated and the assumption that the average rate of mutation will be the same in all regions of the genome and for different organisms.

Once the clock has been calibrated it is then possible to determine when other events in human evolution occurred, for example fossils of our own species, *Homo sapiens*, have been found throughout the old world dating to 100 000 years ago but molecular clock analysis suggests that humans have been around for 150 000 years. Could there still be fossils waiting to be discovered?

[](http://upload.wikimedia.org/wikipedia/commons/6/64/Human-phylo-tree.png)

*Phylogenetic tree showing the evolution of humans (from Wikipedia).*

Molecular clock analysis can be used to build up a phylogenetic tree, as in the above example. A timescale in millions of years is provided along with a genetic distance: this shows that humans and chimpanzees have DNA sequence differences of about 2%.

**Which part of the genome to use?**

Two regions of DNA which are routinely used in this sort of phylogenetic analysis can be found on the male Y chromosome or on the mitochondrial DNA. This allows the molecular clock to be traced through the male or female ancestral lines. (When a sperm fertilises an egg, only the egg’s mitochondria are present in the zygote, so the mitochondria are derived from the mother.)

One advantage of using mitochondrial DNA is the simplicity of a single small chromosome, a haploid genome, which does not have a matching homologous partner with different alleles or crossing over events to complicate the sequence. Mitochondrial chromosomes also have very little non-coding DNA, with genes tightly packed together and with few introns. This is in contrast to normal eukaryote genes, which are usually widely dispersed on a linear chromosome and have numerous introns.