THE RELEVANCE OF MOLECULAR CLOCKS

- Swarbhanu Chatterjee

According to the neutral theory of evolution, for any given DNA sequence, mutations accumulate at an approximately constant rate as long as the DNA sequence retains its original functions. The theory pertains to amino acid substitutions in proteins and nucleotide substitutions in DNA and is therefore concerned with phenomena at the molecular level. It emphasizes that genetic drift is more important in molecular evolution than natural selection. The theory holds that most of the evolution in DNA and protein levels is dominated by drift and, therefore, at the *molecular level,* evolution is mostly non-adaptive and therefore the theory suggests that there is a molecular clock for timing evolutionary events. If the species diverged from a common ancestor, the difference between the mutations of different species would be proportional to time. While this is certainly true for neutral alleles, i.e. those which don't have a phenotype expression, it is to be debated whether the neutrality theory stands ground for those genes which are not neutral. It has been observed recently that for many genes, the variance of the rate of evolution at different times and different lineage is generally larger than what the neutrality theory would lead one to expect and this has raised the question whether the molecular clock is a useful concept at all.

Not only is it true that different genes evolve at different rates in different species, it has recently been found in viruses that the rate of evolution is also dependent on the mode of transmission. In their paper 'Different population dynamics of human T cell lymphotropic virus type II in intravenous drug users compared with endemically infected tribes', Marco Salemi , Martha Lewis, John Fergal Egan, William W. Hall, Jan Desmyter, and Anne-Mieke Vandamme have made the observation that molecular clock analysis show that the HTLV-II virus has two different evolutionary rates in intravenous drug users < IDU > and in endemically infected tribes < the Pygmies and the Amerindians > - it ticks 150-300 times faster in the former. This sharp difference in evolutionary rates seems to be related to the mode of transmission.

Different population dynamics of human T cell lymphotropic virus type II in intravenous drug users compared with endemically infected tribes

The authors used the HLTV-II LTR sequences available in the European Molecular Biology Laboratory and GenBank databases. The LTR is the most variable region in the HLTV-II genome and that along with the fact that it is also the largest dataset available for the fragment corresponding to nucleotides 315-706 of the HLTV-II a Mo isolate is the reason why it was chosen. Two alignments were obtained using the above fragment. One was called II end and it included 26 II a and II b strains isolated from different Amerindian tribes, the Efe2 II d strain isolated from a Mbuti Efe pygmy and the HTLV-II b strain PYGCAM-I isolated from Bakola pygmies. The second one called II idu, included 42 II a and II b IDU strains. All the strains were isolated and sequenced within the last four years. An additional alignment was obtained with 13 II a IDU strains and a II

b IDU strain isolated between 1995 and 1997 and the HTLV-II a Mo strain, which was isolated and sequenced in 1983.

A likelihood mapping analysis was performed which investigated a number of random quartets of sequences. For each quartet, there are three unrooted tree topologies, and their likelihood can be simultaneously represented inside an equilateral triangle as a dot. The corners of the triangle represent the three possible fully resolved tree topologies for the quartet, while the center of the triangle represents a completely unresolved tree topology, or star-like evolution, and the sides reflect network-like evolution. *n*!/4! quartets are possible for n sequences. When *n* is large, a random sample of 10,000 quartets is sufficient to obtain a comprehensive picture of the kind of phylogenetic information present in a particular alignment. The likelihood mapping analysis, with 10,000 random quartets, was performed on the IIend and IIidu alignments, using Tamura-Nei with γ-distributed rates across sites (TNg) implemented in the program PUZZLE as the best nucleotide substitution model.

When the 10,000 random quartets are evaluated with the likelihood-mapping method, one gets different answers for the LTR of JTLV-II in endemically infected populations < such as the pygmy > when compared with HTLV-II infected IDU s. For the former, i.e. the II end alignment, most of the random quartets for the IIend alignment are equally distributed in the three corners of the triangle, representing well-resolved phylogeny, whereas just 14.1% of the quartets are in the center of the triangle, representing star-like phylogeny. The estimated shape parameter = 1.49 ± 1.27 of the γ-distribution results in a bell-shape distribution of the rates along the sites, which means that most sites in the LTR of HTLV-II present in endemically infected populations have intermediate substitution rates and that just a few of them have much lower or much higher rates. In contrast, 45.8% of quartets fall in the center of the triangle when the IIidu alignment is considered, and the estimated shape parameter = 0.22 ± 0.08 of the γ-distribution results in an L-shape distribution of the rates along the sites, which means that most sites in the LTR of HTLV-II isolated in IDUs have very low rates; they are virtually invariable, whereas a few sites exist with very high rates. In the latter case, the standard error in estimating was low, strengthening confidence in the results.

It was found that HTLV-II evolves at a constant rate within IDU s or within endemically infected tribes, irrespective of the subtypes, but at a different rate when one compares between the two populations. Both the likelihood-mapping method and the split decomposition method reveal different evolution dynamics in the two populations. Among endemically infected tribes, HTLV-II evolves in a tree-like manner, whereas among IDUs, there is much more phylogenetic noise, resulting in a star-like evolution. Therefore the branching pattern analysis of the phylogenetic tree shows that the virus linearly increases its population with time among endemically infected tribes while it exponentially grows among IDU s.

On the basis of the above observations, one can say that the evolutionary rate of HTLV-II appears to depend on the mode of transmission. The authors are of the opinion that HTLV-II infected cells in vivo are subjected to clonal expansion and therefore, the

reverse transcription step is less necessary for the virus to maintain its population in the host during its lifetime. On the other hand, it would be more necessary when the virus infects a new host. The higher the transmission rate, the higher is the probability of mutations and hence so is the evolutionary rate of the virus. Furthermore, many HTLV-II positive IDU s are also infected with the HIV virus. It is possible that the presence of HIV and a weaker immune system could affect the HTLV-II evolutionary rate. However, no effect of HIV on the expression and replication cycle of HTLV-II has been ever reported. And so the authors are of the opinion that the dramatic increase in the evolutionary rate within IDUs is probably not related to co-infection with HIV but rather to the increased rate of transmission of HTLV-II.

Motivation

The above results are interesting because they seriously challenge the accuracy of the present day method of estimating the time of evolution using the concept of a 'molecular clock'. In order to have a safe estimate of the time of evolution, one should use carbon dating techniques of fossils in addition to DNA–dating techniques.

Papers

1. 'Different population dynamics of human T cell lymphotropic virus type II in intravenous drug users compared with endemically infected tribes', Marco Salemi, Martha Lewis, John Fergal Egan, William W. Hall, Jan Desmyter, and Anne-Mieke Vandamme, *Proc. Natl. Acad. Sci. USA*, Vol. 96, Issue 23, 13253-13258, November 9, 1999.

2. 'Erratic overdispersion of three molecular clocks: GPDH, SOD, and XDH', Francisco Rodríguez-Trelles, Rosa Tarrío, and Francisco J. Ayala, *Proc. Natl. Acad. Sci. USA*, Vol. 98, Issue 20, 11405-11410, September 25, 2001.