

Poster presentation

Open Access

## Highly improved dating of primate ERV integrations

Hugo Martins<sup>\*1,2</sup> and Palle Villesen<sup>\*1</sup>

Address: <sup>1</sup>Bioinformatics Research Centre, University of Aarhus, Denmark and <sup>2</sup>PhD Program in Computational Biology, Instituto Gulbenkian de Ciências, Oeiras, Portugal

\* Corresponding authors

from *Frontiers of Retrovirology: Complex retroviruses, retroelements and their hosts* Montpellier, France. 21-23 September 2009

Published: 24 September 2009

*Retrovirology* 2009, **6**(Suppl 2):P54 doi:10.1186/1742-4690-6-S2-P54

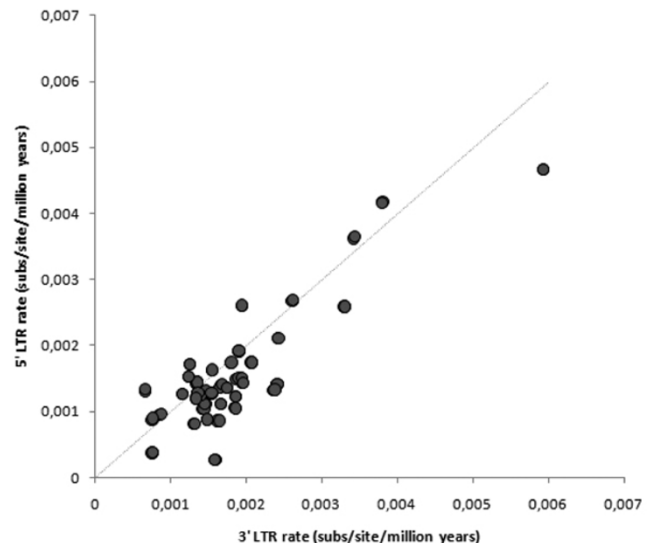
This abstract is available from: <http://www.retrovirology.com/content/6/S2/P54>

© 2009 Martins and Villesen; licensee BioMed Central Ltd.

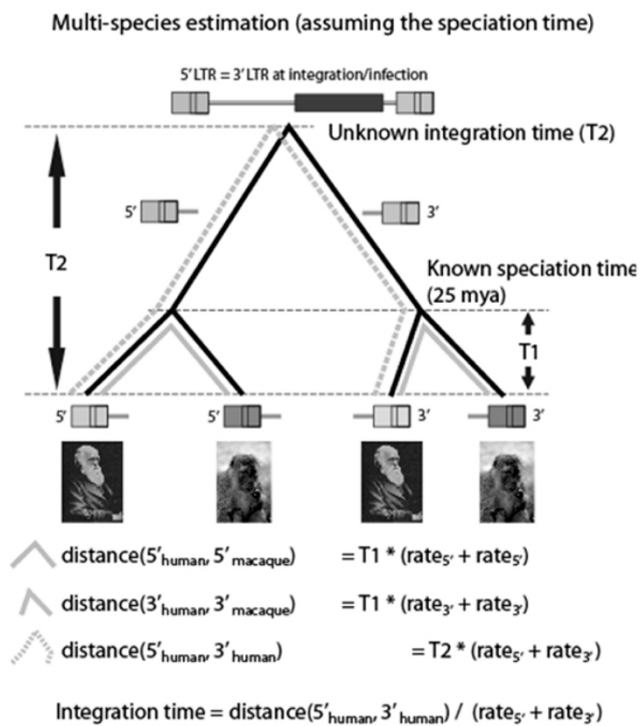
Endogenous retroviruses (ERVs) are genetic fossils of ancient retroviral integrations that remain in the genome of many organisms. Because these remnants are present in many related species, they have become an interesting and useful tool to study phylogenetic relationships [1]. The determination of the insertion time of these viruses has been based upon the assumption that both 5' and 3' Long Terminal Repeats (LTRs) sequences are identical at the time of insertion, but evolve separately afterwards. Similar approaches have been using either a constant evolutionary rate or a range of rates for these viral loci, and only single species data. These methods, however, are based on a very general and wrong assumption: that both LTRs evolve at the same rate [2] (figure 1). Instead, we show that there are strong advantages in using multiple species data and state-of-the-art phylogenetic analysis. We incorporate both simple phylogenetic information and Monte Carlo Markov Chain (MCMC) methods to date the insertions of these viruses based on a relaxed molecular clock approach over a Bayesian phylogeny model and applied them to several selected ERV sequences in primates. These methods treat each ERV locus as having two distinct evolutionary rates for each LTR, and make use of consensual speciation time intervals between primates to calibrate the relaxed molecular clocks (figure 2). Our results show strong improvements when applying simple inference methods that take in account the obtained branch lengths and is computationally inexpensive.

### Conclusion

It is possible to get more robust and realistic integration time estimates by incorporating multiple species data whenever available. A more computationally expensive approach such as the MCMC might be superior but impractical for genome-scale annotations.



**Figure 1**  
Variation between 3' and 5' LTR rates in studied loci as opposed to an uniform rate.



**Figure 2**  
**Estimation of insertion times based on multi-species phylogenetic data.** Known speciation times (T1) can be used in a relaxed molecular clock approach in order to assess the estimated insertion time (T2).

## References

1. Blikstad V, Benachenhou F, Sperber GO, Blomberg J: **Evolution of human endogenous retroviral sequences: a conceptual account.** *Cell Mol Life Sc* 2008, **65(21)**:3348-3365.
2. Rambaut A, Bromham L: **Estimating divergence dates from molecular sequences.** *Mol Biol Evol* 1998, **15**:442-448.