



POSTER PRESENTATION

Open Access

Conservation of ancient full-length open reading frames in vertebrate endogenous retroviruses

Hugo Martins^{1,2*}, Palle Villesen¹

From *Frontiers of Retrovirology* 2011

Amsterdam, The Netherlands. 3-5 October 2011

Background

Endogenous retroviruses (ERVs) are genetic remnants of exogenous retroviral infections that have endured selective processes and made their way into ancestral host genomes. While many of them have deteriorated beyond the ability to code for detectable proteins, some still retain activity or the potential for coding activity [1-3]. A few active ERVs have been coopted into fully functional genes that play important roles in development or disease [1, 2, 4]. New high throughput data revealed that these ERVs are far more common than previously thought and different software packages have been created to search genomes for these structures. In this study, we build on such a tool to look at several vertebrate genomes, search for ERV remnants and identify loci with long, intact open reading frames for the major retroviral genes and dating these loci of interest using the long terminal repeat (LTR) divergence method [5].

Materials and methods

The genomes of twelve vertebrates were acquired from UCSC Genome Browser and analyzed with an in-house method based on the *Itrharvest*[6] software package, complemented with additional filtering and data retrieval. Candidate ERV loci were then probed for full-length open reading frames, which were classified within one of the three major retroviral genes: *gag*, *pol* or *env* by a BlastX [7] against a retroviral protein database. LTR dating for each ERV locus was done by comparing the genetic divergence between the LTRs.

Results/Conclusions

Our study reveals the existence of a few conserved full length open reading frames in loci whose LTRs present a similarity lower than 70% and plenty of open reading

Table 1 Number of ERV loci containing large intact open reading frames (at least 1000 uninterrupted nucleotides) for all three major retroviral genes: *gag*, *pol* and *env*, grouped by their locus/LTR divergence percentage

	<80 %	80-85%	85-90%	90-95%	>95 %	total
Human	0	0	0	1	8	10
Chimpanzee	0	0	0	0	3	3
Orangutan	0	0	0	1	0	1
Macaque	0	0	0	0	2	2
Dog	0	0	0	0	1	1
Mouse	0	0	0	3	169	172
Cow	0	0	0	1	7	8
Opossum	0	0	0	7	33	40
Chicken	0	0	0	0	2	2

frames occur across all genomes in loci with LTR similarity below 80%. Current method limitations may cause this number to be an underestimation but nonetheless reveals the existence of old retroviral infections that have, to this date, kept some coding potential. In the studied primate genomes, a select number of loci still retain long to full length open reading frames on all three major genes even on loci with LTR divergence under 95% (Table 1). *Pol* genes appear to possess the highest number of conserved ORF and a high number of recent integrations with full conservation has been found in the mouse genome.

Author details

¹Bioinformatics Research Centre, Aarhus University, Denmark. ²PhD Program in Computational Biology, Instituto Gulbenkian de Ciências, Oeiras, Portugal.

Published: 3 October 2011

References

1. Simpson GR, et al: Endogenous D-type (HERV-K) related sequences are packaged into retroviral particles in the placenta and possess open reading frames for reverse transcriptase. *Virology* 1996, **222**:451-456.

¹Bioinformatics Research Centre, Aarhus University, Denmark
Full list of author information is available at the end of the article

2. Mi S, et al: Syncytin, a captured retroviral envelope protein involved in human placental morphogenesis. *Nature* 2000, **403**:785-789.
3. Takeuchi Y, et al: Host range and interference studies of three classes of pig endogenous retrovirus. *J Virol* 1998, **72**:9986-9991.
4. Nexø BA, et al: The etiology of multiple sclerosis: genetic evidence for the involvement of the human endogenous retrovirus HERV-Fc1. *PLoS One* 2011, **6**(2):e16652.
5. SanMiguel P, et al: The paleontology of intergene retrotransposons of maize. *Nat Genet* 1998, **20**:43-45.
6. Ellinghaus D, et al: LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinf* 2008, **9**:18.
7. Altschul SF, et al: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997, **25**:3389-3402.

doi:10.1186/1742-4690-8-S2-P47

Cite this article as: Martins and Villesen: Conservation of ancient full-length open reading frames in vertebrate endogenous retroviruses. *Retrovirology* 2011 **8**(Suppl 2):P47.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

