Phylogenetics

**SlidingBayes: exploring recombination using a sliding window approach based on Bayesian phylogenetic inference**

D. Paraskevis\(^1,2,*,\) K. Deforche\(^1\), P. Lemey\(^1\), G. Magiorkinis\(^2\), A. Hatzakis\(^2\) and A.-M. Vandamme\(^1\)

\(^1\)Laboratory for Clinical and Epidemiological Virology, Rega Institute for Medical Research, Katholieke Universiteit Leuven, Minderbroedersstraat 10, B-3000 Leuven, Belgium and \(^2\)National Retrovirus Reference Center, Department of Hygiene and Epidemiology, Athens University Medical School, Mikras Asias 75, GR-11527, Athens, Greece

Received on October 6, 2003; revised on February 26, 2004; accepted on April 20, 2004

Advance Access publication November 16, 2004

**ABSTRACT**

Summary: We developed a software tool (SlidingBayes) for recombination analysis based on Bayesian phylogenetic inference. SlidingBayes provides a powerful approach for detecting potential recombination, especially between highly divergent sequences and complex HIV-1 recombinants for which simpler methods like neighbor joining (NJ) may be less powerful. SlidingBayes guides Markov Chain Monte Carlo (MCMC) sampling performed by MrBayes in a sliding window across the alignment (Bayesian scanning). The tool can be used for nucleotide and amino acid sequences and combines all the modeling possibilities of MrBayes with the ability to plot the posterior probability support for clustering of various combinations of taxa.

Availability: SlidingBayes is available at http://www.kuleuven.ac.be/rega/cev/Software/

Contact: dparask@cc.uoa.gr

Supplementary information: A quick guide and examples for SlidingBayes are available at http://www.kuleuven.ac.be/rega/cev/Software/

**INTRODUCTION**

RNA viruses rate among the most highly variable human pathogens mainly due to the virus’s high substitution and turnover rate, as well as their ability to recombine at high frequency. Recombination has been detected in several viruses such as HIV-1, HBV and Dengue virus (Robertson et al., 1995a,b; Bollyky et al., 1996; Holmes et al., 1999; Worobey et al., 1999), as well as between similar immunodeficiency viruses (SIVs) infecting different primate species (Courgnaud et al., 2003; Salemi et al., 2003; Bailes et al., 2003).

Several strategies have been developed for detecting recombination based on: (1) phylogenetic analysis and sequence comparisons, (2) population genetics methods or (3) site pattern analysis (links and references available at http://bioinf.man.ac.uk/~robertson/recombination/) (Suchard et al., 2002). Among the most frequently used approaches for detecting recombination is phylogenetic analysis using a sliding window. In this approach, bootstrap support for different topological partitions is plotted for a data window that slides along the alignment. This method was first implemented in the bootscanning package (Salminen et al., 1995) and more recently in the Simplot (version 3.2) developed by S. Ray (Lole et al., 1999). The phylogenetic inference method currently implemented in Simplot (version 3.2) is the Neighbor-joining method (NJ) approach under the F84 nucleotide substitution model (Felsenstein, 1984) as the most complex evolutionary model available.

To facilitate exploration of recombination within highly divergent nucleotide or amino acid sequences, we developed a software tool (SlidingBayes) for Bayesian scanning. In this analysis, the approximate posterior probabilities of different partitions are plotted for a sliding data window moving along the alignment.

**SYSTEMS AND METHODS**

Our objective was to develop a tool capable of performing Bayesian phylogenetic analysis using a sliding window approach and displaying the results of all the runs in an easily interpretable way. SlidingBayes uses a graphical user interface (GUI), developed in Java (version 1.4.0) that runs on Windows®, Mac OS X™, and UNIX™ operating systems. SlidingBayes does not estimate the phylogeny itself, but rather interacts with MrBayes v3.0 (Huelsenbeck et al., 2001), which is used for Bayesian phylogenetic inference. SlidingBayes provides the ability to edit the MrBayes command block and thus takes advantage of all available modeling options of MrBayes. For example, user may select from among different evolutionary models and perform phylogenetic analysis on nucleotide or amino acid sequences.

SlidingBayes has two main parts (menus): Run and Analysis. Under the Run menu, the program reads an input file that can contain nucleotide or amino acid sequence alignments and performs the multiple phylogenetic analyses, using a sliding window. More specifically, phylogenetic reconstruction is repeated considering only the data within a window that moves along the alignment in user defined steps. This part of the analysis can be run either within SlidingBayes, or submitted as a batch job to MrBayes; for this purpose, SlidingBayes prepares all necessary input files by breaking down the alignment into overlapping fragments and generating a MrBayes command block. The second part of the program—Analysis menu—reads the output from MrBayes and plots the posterior probability (PP) support in favor of topological partitions consisting of two or more taxa, or in favor of clusters of a user-supplied group of taxa. The former option provides the ability to examine the PP support of every single partition in the dataset and the latter the ability to search for potential recombination between one or more
tarea and a set of ‘reference’ sequences, in a method similar to that in the Simplot program. However, as opposed to Simplot, partitions can be chosen both before and after performing the sliding window analysis. Moreover, by using SlidingBayes, it is possible to open and edit all previous Bayesian scanning, facilitating the analysis for potential recombination between multiple sequences.

It should be noted that the user should choose the length of the Markov chain Monte Carlo (MCMC) chains and the burn-in carefully after the inspection of every single file (window) throughout the alignment. A more detailed description of how to choose the MCMC lengths and burn-in is provided at http://www.kuleuven.ac.be/rega/ccev/Software/in the quick guide for SlidingBayes. Given that the computation time required for Bayesian inference increases with the number of sequences in the input, we suggest limiting the input dataset to as few taxa as possible (<20). However, we note here that there is no limitation in the number of sequences that can be included for SlidingBayes on theoretical grounds.

**ANALYSIS OF HIV-1/SIVcpz AND COMPLEX HIV-1 RECOMBINANT SEQUENCES AS MODEL**

**HIV-1/SIVcpz analysis**

Bayesian scanning analysis was used to explore the phylogenetic relationship between HIV-1 and SIVcpz sequences isolated from Pan troglodytes troglodytes (P.t.t.) (Paraskevis et al., 2003). Using bootscanning analysis, we were not confident in identifying the borders of regions with discordant phylogenies, in a dataset with divergent sequences comprising of HIV-1 and SIVcpz sequences.

On the other hand, the Bayesian scanning method provided a clearer picture of the boundaries of the regions with discordant phylogenetic relationships within HIV-1/SIVcpz than bootscanning analysis, suggesting that, in deep phylogenies, a more sophisticated phylogenetic analysis tool than NJ with F84 may be needed for recombination analysis.

**ANALYSIS OF HIV-1 COMPLEX RECOMBINANTS**

The Bayesian scanning method was also applied in the analysis of complex HIV-1 recombinants in protease (PR) and partial reverse transcriptase (RT) regions (Paraskevis et al., 2005). Some of the HIV-1 recombinants, especially those originating from Central-West Africa, are very complex, consisting of more than two different subtypes in PR/RT partial RT, thus rendering their analysis a difficult task.

We found evidence that in the case of two complex HIV-1 mosaics, originating from Angola (Abecasis et al., 2005), recombination patterns could not be precisely inferred by bootscanning analysis; whereas Bayesian scanning, using SlidingBayes provided a much clearer picture of the potential recombination pattern than the former method.

We should stress here that in both cases—HIV-1/SIVcpz and inter-subtype HIV-1 recombinants—phylogenetic clustering suggested by Bayesian scanning were further confirmed by conditional phylogenetic analysis using both maximum-likelihood (ML) and Bayesian methods (Paraskevis et al., 2003, 2005).

Bayesian scanning is an efficient way to analyze recombination, especially among highly divergent sequences with complex recombiant structues. In these situations, simpler methods for phylogenetic inference like NJ may be inadequate. SlidingBayes is a user-friendly software tool for Bayesian scanning that combines the power of Bayesian inference (Huelsenenbek et al., 2002) with the ability to plot and analyze PP support for different combination of taxa. The program can be run under several different operating systems and it may be used for the recombination exploration or other forms of sliding window analyses using PP support based on nucleotide or amino acid datasets.

**ACKNOWLEDGMENTS**

We wish to acknowledge Marc Suchard for his supportive comments regarding SlidingBayes and for his help in editing this manuscript. P.L. was supported by a Marie Curie fellowship from the European Commission (QLK2-CT2001-51062); P.L. was supported by the Flemish Institute for Scientific-technological Research in Industry (IWT). This work was supported by the Flemish Fonds voor Wetenschappelijk Onderzoek (FWO G.0288.01).

**REFERENCES**


