**Mini Review**

**A Note on Evolutionary Rate Estimation in Bayesian Evolutionary Analysis: Focus on Pathogens**

Kayhan Azadmanesh¹, Sana Eybpoosh² ³ *

¹Department of Virology, Pasteur Institute of Iran, Tehran, Iran;  
²Department of Epidemiology and Biostatistics, Research Centre for Emerging and Reemerging Infectious Diseases, Pasteur Institute of Iran, Tehran, Iran;  
³HIV/STI Surveillance Research Center, and WHO Collaborating Center for HIV Surveillance, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran.

Received May 30, 2017; Accepted Jun 29, 2017

Bayesian evolutionary analysis provide a statistically sound and flexible framework for estimation of evolutionary parameters. In this method, posterior estimates of evolutionary rate ($\mu$) are derived by combining evolutionary information in the data with researcher’s prior knowledge about the true value of $\mu$. Nucleotide sequence samples of fast evolving pathogens that are taken at different points in time carry evolutionary information that allow for estimation of evolutionary rates and divergence dates. If the amount of genetic change in the data is proportional to the time elapsed since divergence from the common ancestor, then one can directly estimate the $\mu$ from the data. Otherwise, external sources should be used to select the $\mu$ value, and use it as a fixed prior in Bayesian evolutionary analysis. This note provides a brief overview on how to assess the adequacy of the evolutionary information in the data and provides some recommendations for obtaining proper evolutionary rate priors from external sources. The recommendations generally highlight the need for the candidate $\mu$ prior to be a good representative of the evolutionary rate in the data at hand. This will be achieved by ensuring that the samples that are the source of the candidate $\mu$ value have been under relatively similar evolutionary forces as the data at hand. As the evolutionary forces acting on a particular set of samples varies across different study settings and species type, selection of prior for $\mu$ should be founded on a thorough understanding of the species under study at biological and social levels. *J Med Microbiol Infec Dis* 2016, 4 (1-2): 8-10.

**Keywords:** Evolution, Evolutionary rate, Bayesian Evolutionary Analysis, Phylogeny.

Recently, there has been growing interest in the analysis of serial nucleotide sequence samples that are taken at different points in time, for fast evolving pathogens [1]. The various sampling dates allow for calibration of evolutionary changes and estimation of absolute evolutionary rates ($\mu$) and divergence times [2]. Bayesian molecular clock dating methods provide a posterior estimate of rate and time. The posterior is derived by combining evolutionary information in serially sampled sequence data with researcher’s prior knowledge about the value of these evolutionary parameters [3].

Generally speaking, for accurate estimation of evolutionary rates, sampled sequences need to cover a wide time span, so that substantial evolution can occur in the data. Also, given a fixed sampling time span, sequences with higher mutation rates, lower heterogeneity, and longer lengths are more suitable for evolutionary rate estimation as opposed to those with lower mutation rate, higher heterogeneity, and shorter sequence length. Moreover, for proper estimation of evolutionary rates, it is desired for the amount of divergence in the dataset to increase linearly with time. Such “clock-like” behavior of the data can be statistically tested using root-to-tip linear regression or maximum likelihood methods. These methods test if the magnitude of genetic divergence of the sequences significantly increases with time (*i.e.*, if $\mu$ and its confidence interval excludes ‘zero.’) For a detailed explanation of these methods see [2]). Root-to-tip regression of genetic distances against sampling time can be done using tools such as TempEst (formerly known as Path-O-Gen) [4]. In cases of data with no clock-like behavior, direct estimation of evolutionary rates from data is not recommended. In such situations, Bayesian estimation of evolutionary rate should be based on evolutionary rate priors obtained from external sources. Two major external sources can be used for this purpose, including 1) external sequence datasets and 2) available literature reporting estimates for $\mu$ [5, 6].

Several studies have estimated evolutionary rates for different pathogen genes. Online genetic databases, such as

---

*Correspondence:* Sana Eybpoosh  
Department of Epidemiology and Biostatistics, Pasteur Institute of Iran, No. 69, Pasteur Ave, Tehran, Iran, 1316943551.  
Email: sana.eybpoosh@gmail.com  
Tel: +98 (21) 64112121 Fax: +98 (21) 66465132

http://jommid.pasteur.ac.ir
selecting an external dataset that includes patients with relatively similar levels of immunity as the ones in the study dataset would also be beneficial. In the circumstances such as HIV infection, the host’s immune system aggravates as the disease progresses. In such cases, selecting an external dataset that comprises patients with relatively similar disease stage as the ones in the study dataset would be preferred.

Finally, different genes and genomic regions are believed to be subjected to various selective pressures and evolutionary constraints (e.g., higher pressure on coding vs. non-coding regions, and on 1\textsuperscript{st} and 2\textsuperscript{nd} codons vs. 3\textsuperscript{rd} codon) [8]. Therefore, it is strongly recommended to pick similar genes or genomic regions from external datasets for evolutionary rate estimation.

This note provides a brief overview of the factors affecting the evolutionary rate of pathogens that need to be considered when obtaining evolutionary rate priors from external sources. The recommendations given here, focus on selecting an external dataset for obtaining evolutionary rate priors. However, these recommendations are also applicable when obtaining the evolutionary rate prior from the existing literature. It is noteworthy that these recommendations may vary depending on the context of the research study, and species under investigation. Selection of external datasets or evolutionary rate estimates should be founded on a thorough understanding of factors affecting the evolutionary rate of species under study at biological and social levels.

ACKNOWLEDGEMENT

The study was not supported by any grants or funds.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest associated with this manuscript.

REFERENCES


