Methods and classes for creating *in silico* evolved genetic sequences of HIV.

Christopher E. Ormsby, Santiago Ávila-Ríos, Gustavo Reyes-Terán*.

Department for Research in Infectious Diseases, National Institute of Respiratory Diseases, Mexico City, Mexico. christopher.ormsby@cieni.org.mx.
*Corresponding author (reyesteran@cieni.org.mx)

Keywords: Simulation, evolution, genomic, phylogeny.

Research on controlling viral infections through vaccines or other strategies relies on methods for measuring the effects of different selective pressures on viruses, and is based on phylogenetic trees which are inferred from the observed data (1).
Simulated genetic evolution has the advantage of being able to know precisely the complete phylogenetic history and all the selective processes that acted on it. Viruses provide an excellent opportunity to use these simulations, since they have a relatively small and simple genome. The data generated can then be used for benchmarking viral evolution models.
We designed a model which accounts for the mayor human immunodeficiency virus genetic selective pressures identified to date, namely HLA driven escape, APOBEC induced hypermutation, recombination, antiretroviral therapy and strong bottleneck events during transmission. Additionally, certain positions are prone to reversion to wild type, and others can cause compensatory mutations. The random mutation rate can be applied with any of the existing nucleotide substitution models. The mutation rate is biased by finite-state reported associations that are applied stochastically at each node. These biases (selective pressures) are read externally from tabular data.
The generated viruses are stored as an object (of novel class `virolver`), with only substitutions from the parent registered, which avoids large memory usage and allows more efficient iterations. Different components of the class define the virus as a founder (i.e. was transmitted from a different subject) or a quasispecies in the individual, HLA subject composition and parent node.
Output is achieved by methods that produce a `phylo` object (used by the CRAN package `ape` (2)), and/or create FASTA format sequences either of the leaves or of all the tips.

References

1. *Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag.* **Carlson JM, Brumme ZL, Rousseau CM, Brumme CJ, Matthews P, Kadie C, Mullins JI,Walker BD, Harrigan PR, Goulder PJ, Heckerman D.** 4, 2008, PLoS Comput Biol, Vol. 11, p. e1000225.

2. **Paradis, E.** *Analysis of Phylogenetics and Evolution with R.* New York : Springer, 2006.