AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology

LETTER TO THE EDITOR

Complexities of Estimating Evolutionary Rates in Viruses

Edward C. Holmes

Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Biological Sciences, and Sydney Medical School, The University of Sydney, Sydney, NSW, Australia

Estimating rates of nucleotide substitution is a powerful way to reveal the dynamics and processes of viral evolution (1). This task has been made easier by the development of sophisticated and user-friendly methods of gene sequence analysis (2). However, the rates estimated are only as reliable as the input data. Indeed, there is a danger that software packages like BEAST (2) are used as a "black box," with poorly constructed data leading to spurious rate estimates and in turn erroneous conclusions about the drivers of viral emergence.

The recent study of Rejmanek et al. (3) falls into this trap. The aim of this work was to measure evolutionary rates in diverse subtypes of influenza A viruses and from this identify the geographic locations where most viral evolution and emergence occurs. To do this, Rejmanek et al. measured rates at the scale of individual countries. However, this analysis is flawed, as many virus sequences sampled from a specific country do not form monophyletic groups. Under these circumstances, viral evolution does not take place within a single country but rather encompasses all the geographic locations occupied by that group of viruses since they last shared a common ancestor. This issue is most acute with human influenza viruses that readily cross geo-political boundaries (4). Hence, it is incorrect to conclude that influenza virus evolves more rapidly in, say, Sweden than in Italy, as in reality both are drawn from the same, geographically mixed, viral population. Although animal influenza viruses often show more structure by country, it is important to check for geographic clustering prior to rate estimation, particularly as the explanations put forward by Rejmanek et al. for why rates differ by location have little connection to the mechanisms that generate genetic diversity.

The evolutionary rates estimated by Rejmanek et al. are also noteworthy for their magnitude; many are higher than observed in other studies of influenza A virus (5, 6), raising the possibility that they are artificially inflated. This might occur in two ways. First, evolutionary rates exhibit time dependency and are often elevated toward the present due to the presence of transient deleterious mutations yet to be removed by purifying selection (7). Second, there may simply be insufficient temporal structure for accurate inference, particularly as many of the sequence data sets were sampled over shallow time ranges. This may explain the wide credible intervals seen for many of the estimates, again illustrating the importance of checking the data prior to rate estimation. In this context, is it also important that the credible intervals be considered when determining whether estimates of substitution rate truly differ, as the mean rates provided give no indication of statistical uncertainty.

Although estimating substitution rates is central to understanding virus evolution, it is essential that these studies be performed with care and employ the necessary quality controls, such as preliminary analyses of phylogenetic and temporal structure. Without such safeguards, it is premature to make predictions about which geographic locations are the key drivers of viral evolution and emergence.

REFERENCES

1. **Duffy S, Shackelton LA, Holmes EC.** 2008. Rates of evolutionary change in viruses: patterns and determinants. Nat Rev Genet **9:**267–276. http://dx.doi.org/10.1038/nrg2323.
2. **Drummond AJ, Suchard MA, Xie D, Rambaut A.** 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol **29:**1969–1973. http://dx.doi.org/10.1093/molbev/mss075.
3. **Rejmanek D, Hosseini PR, Mazet JAK, Daszak P, Goldstein T.** 2015. Evolutionary dynamics and global diversity of influenza A virus. J Virol **89:**10993–11001. http://dx.doi.org/10.1128/JVI.01573-15.
4. **Russell CA, Jones TC, Barr IG, Cox NJ, Garten RJ, Gregory V, Gust ID, Hampson AW, Hay AJ, Hurt AC, de Jong JC, Kelso A, Klimov AI, Kageyama T, Komadina N, Lapedes AS, Lin YP, Mosterin A, Obuchi M, Odagiri T, Osterhaus ADME, Rimmelzwaan GF, Shaw MW, Skepner E, Stohr K, Tashiro M, Fouchier RAM, Smith DJ.** 2008. The global circulation of seasonal influenza A(H3N2) viruses. Science **320:**340–346. http://dx.doi.org/10.1126/science.1154137.
5. **Chen R, Holmes EC.** 2006. Avian influenza virus exhibits rapid evolutionary dynamics. Mol Biol Evol **23:**2336–2341. http://dx.doi.org/10.1093/molbev/msl102.
6. **Worobey M, Han GZ, Rambaut A.** 2014. A synchronized global sweep of the internal genes of modern avian influenza virus. Nature **508:**254–257. http://dx.doi.org/10.1038/nature13016.
7. **Duchene S, Duchene D, Holmes EC, Ho SY.** 2015. The performance of the date-randomization test in phylogenetic analyses of time-structured virus data. Mol Biol Evol **32:**1895–1906. http://dx.doi.org/10.1093/molbev/msv056.

Citation Holmes EC. 2016. Complexities of estimating evolutionary rates in viruses. J Virol 90:2155. doi:10.1128/JVI.02570-15.

Editor: D. S. Lyles

Address correspondence to Edward.holmes@sydney.edu.au.

For the author reply, see doi:10.1128/JVI.02746-15.