

TECHNICAL COMMENT

VIROLOGY

Comment on “Mutation rate and genotype variation of Ebola virus from Mali case sequences”

Andrew Rambaut,^{1,2*} Gytis Dudas,¹ Luiz Max de Carvalho,¹ Daniel J. Park,³ Nathan L. Yozwiak,^{3,4} Edward C. Holmes,⁵ Kristian G. Andersen^{6,7,3*}

Hoenen *et al.* (Reports, 3 April 2015, p. 117; published online 26 March) suggested that the Ebola virus Makona responsible for the West African epidemic evolved more slowly than previously reported. We show that this was based on corrupted data. An erratum provided a rate compatible with the initial and later, more precise, estimates but did not correctly state the nature of the error.

In their Report, Hoenen and colleagues (1) presented an analysis of four Ebola virus (EBOV) genome sequences from Mali in the context of 102 previously published genomes from Guinea and Sierra Leone. Their key assertion was that the evolutionary rate of EBOV during the 2013 to 2016 West African Ebola virus disease (EVD) epidemic was lower than initially reported by Gire *et al.* (2) and similar to the long-term rate of evolution estimated over 35 years in a nonhuman reservoir, presumed to be a bat species. Hoenen *et al.* went on to state that this reported difference in evolutionary rate had important implications for the evolution of transmissibility and virulence of EBOV and for our attempts to control the EVD epidemic. We show that these conclusions are erroneous, based on the corruption of the data by the authors, and paint a false picture of EBOV evolution.

In an erratum to their paper, Hoenen and colleagues (3) substantially revised their evolutionary rate estimates and claimed that this was the consequence of changes to the sample collection dates of the Gire *et al.* (2) data set recorded in GenBank. However, these changes [which only affected 16 of the samples with no change in collection date of more than 6 days (table S1)] do not explain the low rate estimate provided in their original paper (Fig. 1). Hoenen *et al.* kindly

provided their original data file, and we confirm the apparently low evolutionary rate from these data (Fig. 1). However, comparing these sequences with those available on GenBank shows that an extensive shuffling of taxon labels (including date information) among sequences had occurred in their original publication.

This resulted in dates being assigned apparently randomly to sequences, with a subsequent loss of molecular clock signal. This is clearly evident in maximum likelihood phylogenetic trees estimated for both data sets; these possess identical topologies, but there is a clear mixing of label assignments (Fig. 2). As a case in point,

some sequences from the Gire *et al.* (2) data set were from sequential samples from the same patient taken a few days apart. Although these replicate sequences were mostly identical, they often occupy different positions in the Hoenen *et al.* phylogeny [figure 1 in (1)], confirming that a shuffling of sequences must have occurred.

The revised Report states, “By including the newly determined sequences from Mali, we obtained a mean substitution rate of 1.3×10^{-3} substitutions per site per year...This approaches previously reported nucleotide substitution rates of 0.6×10^{-3} to 1.0×10^{-3} for other EBOV sample sets...but is lower than the substitution rate of $\sim 1.9 \times 10^{-3}$ that had been reported for this outbreak.” Although we agree that a rate of 1.3×10^{-3} substitutions per site per year better reflects the evolutionary dynamics of EBOV during this outbreak, this revised estimate is between 1.4 and 2.1 times as high as the between-outbreak estimates cited by Hoenen *et al.* (1). However, the statistical credible interval for the revised Hoenen *et al.* (3) rate estimates is broad (Fig. 1), necessarily reflecting the limited time span and data at that point. The rate reported by Gire *et al.* (1.9×10^{-3}) was affected by the use of the three original Guinea sequences (4), which contained a number of imputation errors, later corrected (figure S2). These erroneous nucleotide sites were corrected for most of the analyses (as described in Gire *et al.*), but not for the rate value reported, and using the revised Guinea sequences brings the rate estimate for the Gire *et al.* data to 1.5×10^{-3} (figure S3). Importantly, both the revised Hoenen *et al.* (3) and Gire *et al.* (2) rate credible intervals are consistent with estimates reported by studies undertaken later in the epidemic (5–7) (Fig. 1).

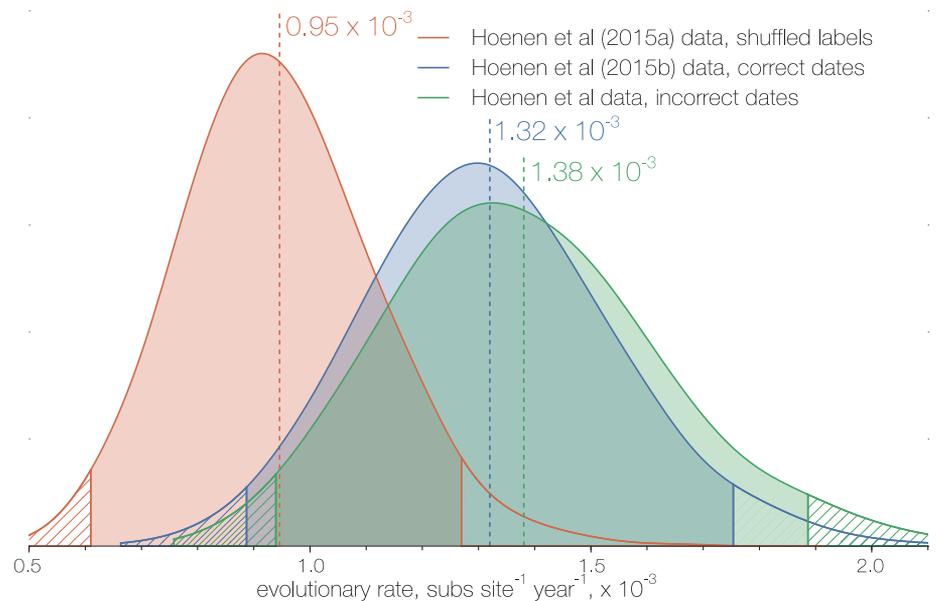


Fig. 1. Posterior probability densities of the mean evolutionary rate estimated for the sequence data provided by Hoenen *et al.*, including both sample shuffling and date errors, as originally published (red), sample date errors only (green), and as fully corrected in the erratum (blue). The vertical lines mark the 95% highest posterior density intervals.

¹Institute of Evolutionary Biology, University of Edinburgh, EH9 3FL, Edinburgh, UK. ²Centre for Immunology, Infection, and Evolution at the University of Edinburgh, EH9 3FL, Edinburgh, UK. ³Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA. ⁴Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. ⁵Marie Bashir Institute of Infectious Diseases and Biosecurity, Charles Perkins Centre, Sydney Medical School, and School of Life and Environmental Sciences, University of Sydney, Sydney, NSW, 2006, Australia. ⁶The Scripps Research Institute, Department of Immunology and Microbial Science, La Jolla, CA 92037, USA. ⁷Scripps Translational Science Institute, La Jolla, CA 92037, USA. *Corresponding author. Email: a.rambaut@ed.ac.uk (A.R.); kga1978@gmail.com (K.G.A.)

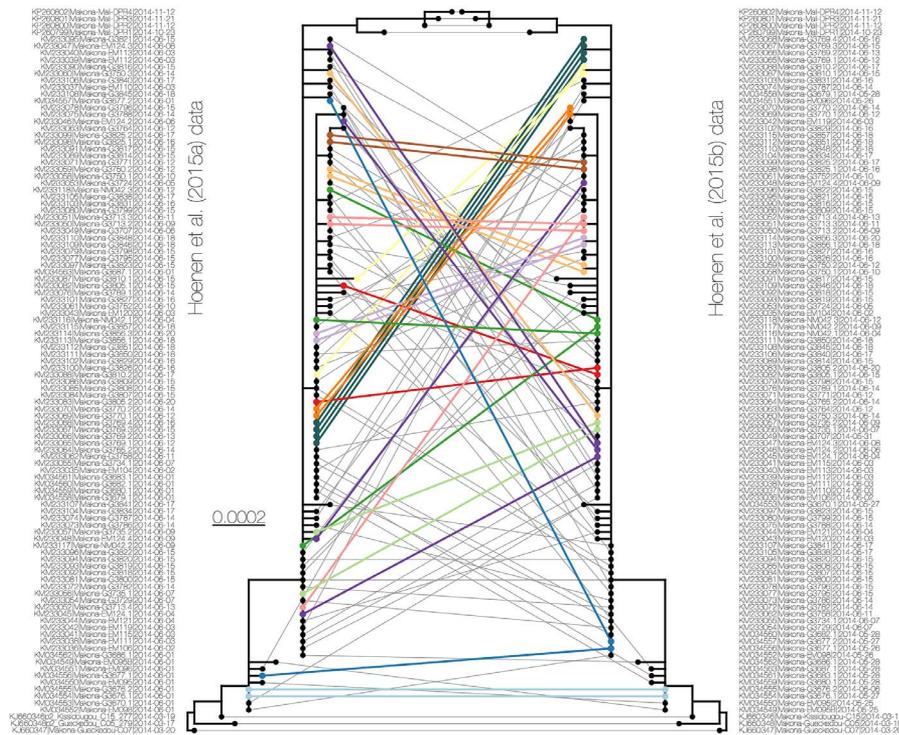


Fig. 2. Maximum likelihood tree of the 106 sequences analyzed by Hoenen *et al.* (left side) initially (1) with lines linking to the correctly labeled sequences after the erratum (3) (right side). Lines of the same color represent multiple samples taken from the same patient, which in most cases have identical sequences. These correctly group together on the right but do not in many cases on the left.

Based on their observed evolutionary rate estimate, Hoenen *et al.* (1) conclude that it is unlikely that the types of genetic changes observed thus far

would impair diagnostic measures or affect the efficacy of vaccines or potential virus-specific treatments. It is overly simplistic to conclude that the

differences in evolutionary rate estimates reported for EBOV will greatly alter its potential to change its virulence and/or transmissibility or our attempts to control it. For example, human dengue virus has been able to generate considerable antigenic diversity (8) that hinders successful vaccination, despite experiencing evolutionary rates usually lower than EBOV Makona (9). In addition, these reported differences in rate estimates are minor compared to the range of evolutionary rates seen in RNA viruses that span approximately 10^{-5} to 10^{-2} substitutions per site per year (9, 10). As underlying mutation rates in EBOV are still likely to be on the order of one mutation per genome replication, genotype variation is undoubtedly generated at a rate sufficient to enable rapid phenotypic evolution should a suitable selection pressure arise (11).

REFERENCES

1. T. Hoenen *et al.*, *Science* **348**, 117–119 (2015).
2. S. K. Gire *et al.*, *Science* **345**, 1369–1372 (2014).
3. T. Hoenen *et al.*, *Science* **348**, aac5674 (2015).
4. S. Baize *et al.*, *N. Engl. J. Med.* **371**, 1418–1425 (2014).
5. Y.-G. Tong *et al.*, *Nature* **524**, 93–96 (2015).
6. M. W. Carroll *et al.*, *Nature* **524**, 97–101 (2015).
7. D. J. Park *et al.*, *Cell* **161**, 1516–1526 (2015).
8. L. C. Katzelnick *et al.*, *Science* **349**, 1338–1343 (2015).
9. R. Sanjuán, *PLOS Pathog.* **8**, e1002685 (2012).
10. S. Duffy, L. A. Shackelton, E. C. Holmes, *Nat. Rev. Genet.* **9**, 267–276 (2008).
11. J. R. Kugelman *et al.*, *Cell Rep.* **12**, 2111–2120 (2015).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/353/6300/aaf3823/suppl/DC1
Figs. S1 to S3
Table S1
References

5 February 2016; accepted 21 June 2016
10.1126/science.aaf3823

Comment on "Mutation rate and genotype variation of Ebola virus from Mali case sequences"

Andrew Rambaut, Gytis Dudas, Luiz Max de Carvalho, Daniel J. Park, Nathan L. Yozwiak, Edward C. Holmes and Kristian G. Andersen

Science **353** (6300), 658.
DOI: 10.1126/science.aaf3823

ARTICLE TOOLS

<http://science.sciencemag.org/content/353/6300/658.1>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2016/08/11/353.6300.658-a.DC1>

RELATED CONTENT

<http://science.sciencemag.org/content/sci/353/6300/658.2.full>
<http://science.sciencemag.org/content/sci/348/6230/117.full>
<http://stm.sciencemag.org/content/scitransmed/8/329/329ra33.full>
<http://stm.sciencemag.org/content/scitransmed/7/317/317ps24.full>
<http://stm.sciencemag.org/content/scitransmed/7/307/307rv5.full>
<http://stm.sciencemag.org/content/scitransmed/7/290/290ra89.full>

REFERENCES

This article cites 11 articles, 4 of which you can access for free
<http://science.sciencemag.org/content/353/6300/658.1#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)