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

T. Hoener,1* D. Safronetz,1* A. Groshet,1* K. R. Wollenberg,2* O. A. Koia,2 B. Diarrha,3 I. S. Fall,4 F. C. Haidara,5 F. Diallo,5 M. Sanogo,3 Y. S. Sarro,3 A. Kone,3 A. C. G. Togo,3 A. Traore,3 M. Kodio,5 A. Dosseh,6 K. Rosenke,1 E. de Wit,1 F. Feldmann,7 H. Ebihara,1 V. J. Munster,1 K. C. Zoon,5 H. Feldmann,†† S. Sow††

1Laboratory of Virology, Division of Intramural Research, National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Hamilton, MT 59840, USA. 2Bioinformatics and Computational Biosciences Branch, NIAID, NIH, Bethesda, MD 20892, USA. 3Center of Research and Training for HIV and Tuberculosis, University of Science, Technique and Technologies of Bamako, Mali. 4World Health Organization Office, Bamako, Mali. 5Centre des Opérations d’Urgence, Centre pour le Développement des Vaccins (CVD-Mali), Centre National d’Appui à la lutte contre la Maladie, Ministère de la Santé et de l’Hygiène Publique, Bamako, Mali. 6World Health Organization Inter-Country Support Team, Ouagadougou, Burkina Faso. 7Rocky Mountain Veterinary Branch, Division of Intramural Research, NIAID, NIH, Hamilton, MT 59840, USA. 8Office of the Scientific Director, NIAID, NIH, Bethesda, MD 20895, USA.

*These authors contributed equally to this work.
††These authors contributed equally to this work.
‡‡Corresponding author. E-mail: feldmannh@niaid.nih.gov (H.F.); ssow@medicine.umd.edu (S.S.)

The occurrence of Ebola virus (EBOV) in West Africa during 2013–2015 is unprecedented. Early reports suggested that in this outbreak EBOV is mutating twice as fast as previously observed, which indicates the potential for changes in transmissibility and virulence and could render current molecular diagnostics and countermeasures ineffective. We have determined additional full-length sequences from two clusters of imported EBOV infections into Mali, and we show that the nucleotide substitution rate (9.6 × 10⁴ substitutions per site per year) is consistent with rates observed in Central African outbreaks. In addition, overall variation among all genotypes observed remains low. Thus, our data indicate that EBOV is not undergoing rapid evolution in humans during the current outbreak. This finding has important implications for outbreak response and public health decisions and should alleviate several previously raised concerns.

In December 2013, an outbreak of Ebola virus (EBOV) started in Guinea with a single index case, resulting in widespread human-to-human transmission in this country, as well as in neighboring regions of Sierra Leone and Liberia. Despite being by far the largest and longest lasting outbreak, there has been limited information regarding the evolution of EBOV. To date, the only sequences published have been from virus isolates derived from three patient samples in Guinea from March 2014 (1) and from a cluster of sequences derived from samples from Sierra Leone from June 2014 (2). However, no new information has been available during the intervening 6 months of large-scale virus circulation, and thus these virus sequences may no longer adequately inform us about the nature of currently circulating strains. This is of particular importance because diagnostics are based predominantly on reverse transcription polymerase chain reaction (3) and are thus sequence-dependent, as are some of the therapeutic options currently being considered for deployment (i.e., small interfering RNA-based treatments, such as TKM-Ebola). Further, mutations in the glycoprotein (GP) could affect the efficacy of vaccines or antibody treatments (e.g., ZMapp). Analyses based on the limited available sequence information have also been used to suggest that EBOV is mutating more rapidly during this outbreak than during previous outbreaks (2), potentially as a result of sustained and large-scale human-to-human transmission, which has, in turn, raised concerns about increased virulence or transmissibility (4).

Although the outbreak has been mostly limited to Guinea, Sierra Leone, and Liberia until now, there have been imported cases in Nigeria, Senegal, Spain, and the United States, without substantial subsequent spread of the virus. Recently, two independent introductions of EBOV into Mali also occurred: The first introduction in October 2014, from Kissidougou, Guinea, resulted in a single fatal case with no further transmission; the second introduction in November 2014, from Kouremale-Guinee, Guinea, resulted in six primary transmissions (five fatal, one nonfatal) and a single secondary transmission (nonfatal). We obtained patient samples from both introductions and determined the full-length sequences of the corresponding EBOV genomes [see (5) for materials and methods]. Surprisingly, the virus from the first introduction (hereafter referred to as “Mali-DPR1,” sampled 23 October 2014) showed only nine nucleotide differences from the most closely related virus (Makona-EM106, accession number KM233036), which was sampled in Sierra Leone almost 6 months earlier (2 June 2014). Additionally, Mali-DPR1 showed only 15 nucleotide differences to the most distantly related EBOV from the West African outbreak (Gueckedou-C05, accession number KJ660348.2) (fig. S1). In terms of amino acids, only one change had been acquired in comparison to the Makona-EM106 virus [g16514a:S=N (6) at position 1645 in the polymerase (L) gene] (fig. S1). The viruses isolated from the second introduction into Mali (hereafter referred to as “Mali-DPR2,” 13
and -4,” sampled 12 to 21 November) were closely related to Mali-DPR1, although they remained clearly distinct from that sequence, with seven to nine nucleotide differences. However, in terms of amino acid sequence, only a single change (t15560c:F>Sat position 1327 in the L gene) was observed in Mali-DPR4, and no amino acid changes were observed in Mali-DPR2 and -3.

Phylogenetic analysis (5) of all published sequences showed that the West African sequences group into several well-supported clades that are consistently identified using Bayesian analysis (Fig. 1), as well as maximum-likelihood and neighbor-joining analyses (fig. S2), producing trees with comparable internal structure. The Malian sequences all form a discrete and well-supported branch derived from the second cluster of Sierra Leone sequences, despite being derived from two independent introduction events originating ~400 km from each other. This result is surprising, given the genetic diversity that had been shown to exist among viruses within the outbreak region during the sampling in Sierra Leone. Although we cannot completely exclude serendipitous introduction of two such closely related isolates, the strong support for coalescence of the Malian viruses to a common ancestor suggests that this genotype may be highly prevalent in this region at present and may thus represent a reasonable proxy for testing the efficacy of diagnostic, therapeutic, and vaccination approaches.

Based on these additional data, we recalculated the nucleotide substitution rate for the West African outbreak (see (5) for a full list of virus genomes). By including the newly determined sequences from Mali, we obtained a mean substitution rate of 9.6 × 10⁻⁴ substitutions per site per year (Fig. 2). This matches previously reported nucleotide substitution rates of 6.2 × 10⁻⁴ to 9.5 × 10⁻⁴ for other EBOV sample sets (7–9) but differs from the substitution rate of ~1.9 × 10⁻³ that had been reported for this outbreak (2). Further analysis of this expanded data set also generated a comparable value (6.9 × 10⁻⁴ substitutions per site per year) when analysis parameters previously used to obtain these higher estimates (2) were employed. Surprisingly, reanalysis of the previously available data set from Gire et al. (2) using the model developed for this study also produced a similar value (8.2 × 10⁻⁴ substitutions per site per year). These findings appear to suggest that the strict clock used in the current analysis may be more suitable and robust for this data set and that the expanded time frame over which samples in our data sets were obtained adds additional robustness to the analysis.

Detailed analysis of the individual mutations observed among all currently known genotypes showed that the majority of mutations were synonymous or occurred in noncoding regions [particularly at the virion protein 30 (VP30)/VP24 gene border, which contains a long intergenic region] (fig. S3). Although up to four nonsynonymous nucleotide changes from the consensus sequence were observed for early sequences (i.e., those from March 2014), almost no nonsynonymous changes were observed in later sequences. Most nucleotide changes occurred only sporadically in a single genotype; however, a few changes were observed in several genotypes and might represent changes that remained conserved in later viruses. Early changes included two nonsynonymous changes, one each in the nucleo- protein (t800c:C>R) and GP (t6283c:C>V); three synonymous changes, one in VP30 (c8928a:P>P) and two in L (a15963g:C>K, c17142t:F>F); and one change in the VP24 noncoding region (g10219a). Later changes conserved among the Malian sequences were one nonsynonymous change in L (g16514a:S=N); four synonymous changes in GP (a6056c:C1>G); VP30 (c8928a:P>P), and L (c14253t:T>G, a15963g:C>K); and an additional change in the VP24 noncoding region (c10315t). However, none of these nonsynonymous changes coincide with known functional domains or motifs within the affected proteins, and overall the genotypes observed have remained stable since June 2014, which is consistent with reports from earlier outbreaks of both EBOV and Marburg virus (10, 11).

In the past, EBOVs have been reported to undergo only limited genetic changes during outbreaks (10, 11), a phenomenon that also seems to be true in the current outbreak, despite prolonged human-to-human transmission. Thus, the potential for acquisition of virulence or increased transmissibility of EBOVs is constrained. Although the size and nature of the mutational targets associated with the acquisition of virulence clearly vary between viruses, and there are cases in which just a few mutations can substantially affect virus properties, there are also many limitations. These include the need for multiple mutations in some cases, as for mammalian transmission of H5N1 influenza (12), and these mutations may need to occur simultaneously or in a defined order (13). Studies aimed at identifying the virulence factors of EBOVs already indicate that virulence is a complex, multifactorial trait for these viruses, although further studies are required to better define these factors (14, 15). Furthermore, epidemiological and case management data do not support increased virulence in humans during the current outbreak (16). Thus, whereas from a public health perspective the current EBOV outbreak in West Africa continues to be an extremely pressing emergency, it is doubtful that either virulence or transmissibility has increased in the circulating EBOV strains. We have also shown that, despite the extensive and prolonged human-to-human transmission in this outbreak, the virus is not mutating at a rate beyond what is expected. Moreover, although it is certainly possible for different virus lineages to exhibit different evolutionary rates, those predicted using samples from various Ebola outbreaks have remained consistent. Similarly, on the basis of our data, 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. Nevertheless, monitoring of the situation remains paramount to ensure that this contin-
uses to be the case as this outbreak progresses and if further outbreaks arise.

REFERENCES AND NOTES


5. Materials and methods are available as supplementary materials on Science Online.

6. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.


ACKNOWLEDGMENTS

We thank S. Nichol and U. Stohrer (U.S. Centers for Disease Control and Prevention); T. Schwab, R. Sakai, M. Niang, and M. Pineda (NIAMID); and Diop, S. Tounkara, and F. Dao (Centre de Recherche et de Formation sur le VIH/EBERF) for helpful discussion and help with logistics and sample preparation. We also thank the Research Technology Branch at the Rocky Mountain Laboratories, particularly S. Kramer, K. Barbarin, and S. Porcella, for sequencing services. This study made use of the high-performance computational capabilities of the Biowulf Linux cluster at the NIH, Bethesda, MD (http://biowulf.nih.gov), and the Office of Cyber Infrastructure and Computational Biology High-Performance Computing cluster at the NIAMD, Bethesda, MD. This work was supported in part by the Intramural Research Program of the NIH, NIAMD. The full-length sequences of the Malian EBOVs were deposited at GenBank under the accession numbers KP260799.1, KP260800.1, KP260801.1, and KP260802.1 (Ebola virus/H.sapiens-wt/MLU/2014/Makona-Mali-DPR1 to 4). There are no conflicts of interest, material transfer agreements, patents, or patent applications that apply to reagents, methods, or data in the paper for any of the authors.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/cgi/content/full/science.aa5646/DC1
Materials and Methods
Figs. S1 to S3
References (S7–23)

24 December 2014; accepted 23 February 2015
Published online 26 March 2015
10.1126/science.aa5646
Fig. 1. Phylogenetic relationship among viruses from the ongoing outbreak. A Bayesian tree of all currently published sequences from the ongoing outbreak is shown. Branch colors indicate posterior probability, with terminal branches shown in black. The x axis indicates time in years before acquisition of the last sample (21 November 2014). Although clade structure in the analysis is ambivalent for some of the Sierra Leone viruses, there is strong support for the placement of the Malian viruses as a distinct lineage originating from the main group of Sierra Leone sequences and distinct from both the Guinean viruses, as well as another well-supported clade of Sierra Leone viruses.
Fig. 2. Nucleotide substitution rate in the current outbreak. The probability distribution of substitution rates in the current EBOV outbreak in West Africa based on all currently published sequences from this outbreak, including those from Mali, is shown as a gray shaded area. Previously reported substitution rates for EBOVs from four different publications (2, 7–9) are indicated by dotted lines (further qualifying information is provided in brackets if several rates were reported in those publications).