Sciencexpress Reports

Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak

Stephen K. Gire,[1,2*] Augustine Goba,[3*†] Kristian G. Andersen,[1,2*†] Rachel S. G. Sealfon,[2,4*] Daniel J. Park,[2*] Lansana Kanneh,[3] Simbirie Jalloh,[3] Mambu Momoh,[3,5] Mohamed Fullah,[3,5‡] Gytis Dudas,[6] Shirlee Wohl,[1,2,7] Lina M. Moses,[8] Nathan L. Yozwiak,[1,2] Sarah Winnicki,[1,2] Christian B. Matranga,[2] Christine M. Malboeuf,[2] James Qu,[2] Adrianne D. Gladden,[2] Stephen F. Schaffner,[1,2] Xiao Yang,[2] Pan-Pan Jiang,[1,2] Mahan Nekoui,[1,2] Andres Colubri,[1] Moinya Ruth Coomber,[3] Mbalu Fonnie,[3‡] Alex Moigboi,[3‡] Michael Gbakie,[3] Fatima K. Kamara,[3] Veronica Tucker,[3] Edwin Konuwa,[3] Sidiki Saffa,[3] Josephine Sellu,[3] Abdul Azziz Jalloh,[3] Alice Kovoma,[3‡] James Koninga,[3] Ibrahim Mustapha,[3] Kandeh Kargbo,[3] Momoh Foday,[3] Mohamed Yillah,[3] Franklyn Kanneh,[3] Willie Robert,[3] James L. B. Massally,[3] Sinéad B. Chapman,[2] James Bochicchio,[2] Cheryl Murphy,[2] Chad Nusbaum,[2] Sarah Young,[2] Bruce W. Birren,[2] Donald S. Grant,[3] John S. Scheiffelin,[8] Eric S. Lander,[2,7,9] Christian Happi,[10] Sahr M. Gevao,[11] Andreas Gnirke,[2§] Andrew Rambaut,[6,12,13§] Robert F. Garry,[8§] S. Humarr Khan,[3‡§] Pardis C. Sabeti[1,2†§]

[1]Center for Systems Biology, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. [2]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [3]Kenema Government Hospital, Kenema, Sierra Leone. [4]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [5]Eastern Polytechnic College, Kenema, Sierra Leone. [6]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK. [7]Systems Biology, Harvard Medical School, Boston, MA 02115, USA. [8]Tulane University Medical Center, New Orleans, LA 70112, USA. [9]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [10]Redeemer's University, Ogun State, Nigeria. [11]University of Sierra Leone, Freetown, Sierra Leone. [12]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA. [13]Centre for Immunity, Infection and Evolution, University of Edinburgh, Edinburgh EH9 3JT, UK.

*These authors contributed equally to this work.

†Corresponding author. E-mail: andersen@broadinstitute.org (K.G.A.); augstgoba@yahoo.com (A.G.); psabeti@oeb.harvard.edu (P.C.S.)

‡Deceased.

§These authors jointly supervised this work.

In its largest outbreak, Ebola virus disease is spreading through Guinea, Liberia, Sierra Leone, and Nigeria. We sequenced 99 Ebola virus genomes from 78 patients in Sierra Leone to ~2,000x coverage. We observed a rapid accumulation of interhost and intrahost genetic variation, allowing us to characterize patterns of viral transmission over the initial weeks of the epidemic. This West African variant likely diverged from Middle African lineages ~2004, crossed from Guinea to Sierra Leone in May 2014, and has exhibited sustained human-to-human transmission subsequently, with no evidence of additional zoonotic sources. Since many of the mutations alter protein sequences and other biologically meaningful targets, they should be monitored for impact on diagnostics, vaccines, and therapies critical to outbreak response.

Ebola virus (EBOV; formerly Zaire ebolavirus), one of five ebolaviruses, is a lethal human pathogen, causing Ebola virus disease (EVD) with an average case fatality rate of 78% (*1*). Previous EVD outbreaks were confined to remote regions of Middle Africa; the largest, in 1976, had 318 cases (*2*) (Fig. 1A). The current outbreak started in February 2014 in Guinea, West Africa (*3*) and spread into Liberia in March, Sierra Leone in May, and Nigeria in late July. It is the largest known EVD outbreak and is expanding exponentially with a doubling period of 34.8 days (Fig. 1B). As of August 19th, 2,240 cases and 1,229 deaths have been documented (*4*, *5*). Its emergence in the major cities of Conakry (Guinea), Freetown (Sierra Leone), Monrovia (Liberia), and Lagos (Nigeria) raises the specter of increasing local and international dissemination.

In an ongoing public health crisis, where accurate and timely information is crucial, new genomic technologies can provide near real-time insights into the pathogen's origin, transmission dynamics, and evolution. We used massively parallel viral sequencing to understand how and when EBOV entered human populations in the 2014 West African outbreak, whether the outbreak is continuing to be fed by new transmissions from its natural reservoir, and how the virus changed, both before and after its recent jump to humans.

In March 2014, Kenema Government Hospital (KGH) established EBOV surveillance in Kenema, Sierra Leone, near the origin of the 2014 outbreak (Fig. 1C and fig. S1) (*6*). Following standards for field-based tests in previous (*7*) and current (*3*) outbreaks, KGH performed conventional PCR-based EBOV diagnostics (*8*) (fig. S2); all tests were negative through early May. On May 25, KGH scientists confirmed the first case of EVD in Sierra Leone. Investigation by the Ministry of Health and Sanitation (MoHS) uncovered an epidemiological link between this case and the burial of a traditional healer who had treated EVD patients in Guinea. Tracing led to 13 additional cases—all females who attended the burial. We obtained ethical approval from MoHS, the Sierra Leone Ethics and Scientific Review Committee, and our U.S. institutions to sequence patient samples in the U.S. using approved safety standards (*6*).

We evaluated four independent library preparation methods and two sequencing platforms (*9*) (table S1) for our first batch of 15 inactivated EVD samples from 12 patients. Nextera library construction and Illumina sequencing provided the most complete genome assembly and reliable intrahost single nucleotide variant (iSNV, frequency >0.5%) identification (*6*). We used this combination for a second batch of 84 samples from 66 additional patients, performing two independent replicates from each sample (Fig. 1D). We also sequenced 35 samples from suspected EVD cases that tested negative for EBOV; genomic analysis identified other known pathogens, including Lassa virus, HIV-1, enterovirus A and malaria parasites (fig. S3).

In total, we generated 99 EBOV genome sequences from 78 confirmed EVD patients, representing over 70% of the EVD patients diagnosed in Sierra Leone in late May to mid June; we employed multiple

extraction methods or timepoints for 13 patients (table S2). Median coverage was >2,000x, spanning more than 99.9% of EBOV coding regions (Fig. 1, D and E, and table S2).

We combined the 78 Sierra Leonean sequences with 3 published Guinean samples (*3*) (correcting 21 likely sequencing errors in the latter (*6*)) to obtain a dataset of 81 sequences. They reveal 341 fixed substitutions between the 2014 EBOV and all previously published EBOV (35 nonsynonymous, 173 synonymous, 133 noncoding), with an additional 55 single nucleotide polymorphisms (SNPs) (fixed within individual patients) within the West African outbreak (15 nonsynonymous, 25 synonymous, 15 noncoding). Notably, the Sierra Leonean genomes differ from PCR probes for five separate assays used for EBOV and pan-filovirus diagnostics (table S3).

Deep-sequence coverage allowed identification of 263 iSNVs (73 nonsynonymous, 108 synonymous, 70 noncoding, and 12 frameshift) in the Sierra Leone patients (*6*). For all patients with multiple time points, consensus sequences were identical and iSNV frequencies remained stable (fig. S4). One notable intrahost variation is the RNA editing site of the glycoprotein (GP) gene (fig. S5A) (*10–12*), which we characterize in patients (*6*).

Phylogenetic comparison to all 20 genomes from earlier outbreaks suggests the 2014 West African virus likely spread from Middle Africa within the last decade. Rooting the phylogeny using divergence to other ebolavirus genomes is problematic (Fig. 2A and fig. S6) (*6*, *13*). However, rooting the tree on the oldest outbreak reveals a strong correlation between sample date and root-to-tip distance, with a substitution rate of $8x10^{-4}$/site/year (Fig. 2B and fig. S7) (*13*). This suggests that the lineages of the three most recent outbreaks all diverged from a common ancestor at roughly the same time c. 2004 (Fig. 2C and Fig. 3A), supporting the hypothesis that each outbreak represents an independent zoonotic event from the same genetically diverse viral population in its natural reservoir.

Genetic similarity across the sequenced 2014 samples suggests a single transmission from the natural reservoir, followed by human-to-human transmission during the outbreak. Molecular dating places the common ancestor of all sequenced Guinea and Sierra Leone lineages around late February 2014 (Fig. 3B), three months after the earliest suspected cases in Guinea (*3*); this coalescence would be unlikely had there been multiple transmissions from the natural reservoir. Thus, in contrast to some previous EVD outbreaks (*14*), continued human-reservoir exposure is unlikely to have contributed to the growth of this epidemic in areas represented by available sequence data.

Our data suggest the Sierra Leone outbreak stemmed from the introduction of two genetically distinct viruses from Guinea around the same time. Samples from 12 of the first EVD patients in Sierra Leone, all believed to have attended the funeral of an EVD case from Guinea, fall into two distinct clusters (clusters 1 and 2) (Fig. 4A and fig. S8). Molecular dating places the divergence of these two lineages in late April (Fig. 3B), pre-dating their co-appearance in Sierra Leone in late May (Fig. 4B), suggesting the funeral attendees were most likely infected by two lineages then circulating in Guinea, possibly at the funeral (fig. S9). All subsequent diversity in Sierra Leone accumulated on the background of those two lineages (Fig. 4A), consistent with epidemiological information from tracing contacts.

Patterns in observed intrahost and interhost variation provide important insights about transmission and epidemiology. Groups of patients with identical viruses or with shared intrahost variation show temporal patterns suggesting transmission links (fig. S10). One iSNV (position 10,218) shared by twelve patients is later observed as fixed within 38 patients, becoming the majority allele in the population (Fig. 4C) and defining a third Sierra Leone cluster (Fig. 4, A and D, and fig. S8). Repeated propagation at intermediate frequency suggests that transmission of multiple viral haplotypes may be common. Geographic, temporal, and epidemiological metadata supports the transmission clustering inferred from genetic data (Fig. 4, D and E, and fig. S11) (*6*).

The observed substitution rate is roughly twice as high within the 2014 outbreak as between outbreaks (Fig. 4F). Mutations are also more frequently nonsynonymous during the outbreak (Fig. 4G). Similar findings have been seen previously (*15*) and are consistent with expectations from incomplete purifying selection (*16–18*). Determining whether individual mutations are deleterious, or even adaptive, would require functional analysis; however, the rate of nonsynonymous mutations suggests that continued progression of this epidemic could afford an opportunity for viral adaptation (Fig. 4H), underscoring the need for rapid containment.

As in every EVD outbreak, the 2014 EBOV variant carries a number of genetic changes distinct to this lineage; our data do not address whether these differences are related to the severity of the outbreak. However, the catalog of 395 mutations, including 50 fixed nonsynonymous changes with 8 at positions with high levels of conservation across ebolaviruses, provide a starting point for such studies (table S4).

To aid in relief efforts and facilitate rapid global research, we immediately released all sequence data as generated. Ongoing epidemiological and genomic surveillance is imperative to identify viral determinants of transmission dynamics, monitor viral changes and adaptation, ensure accurate diagnosis, guide research on therapeutic targets, and refine public-health strategies. It is our hope that this work will aid the multidisciplinary, international efforts to understand and contain this expanding epidemic.

In memoriam: Tragically, five co-authors, who contributed greatly to public health and research efforts in Sierra Leone, contracted EVD in the course of their work and lost their battle with the disease before this manuscript could be published. We wish to honor their memory.

References and Notes

1. J. H. Kuhn, L. E. Dodd, V. Wahl-Jensen, S. R. Radoshitzky, S. Bavari, P. B. Jahrling, Evaluation of perceived threat differences posed by filovirus variants. *Biosecur. Bioterror.* **9**, 361–371 (2011). Medline doi:10.1089/bsp.2011.0051
2. J. Burke, Ebola haemorrhagic fever in Zaire, 1976. *Bull. World Health Organ.* **56**, 271–293 (1978). Medline
3. S. Baize *et al*., Emergence of Zaire Ebola virus disease in Guinea—Preliminary report. *N. Engl. J. Med.* 10.1056/NEJMoa1404505 (2014). doi:10.1056/NEJMoa1404505
4. WHO, (2014), www.who.int/csr/don/archive/disease/ebola/en/
5. O. Reynard, V. Volchkov, C. Peyrefitte, Une première épidémie de fièvre à virus Ebola en Afrique de l'Ouest. *Med. Sci.* **30**, 671–673 (2014). doi:10.1051/medsci/20143006018
6. See supplementary materials on *Science* Online.
7. J. S. Towner, T. K. Sealy, T. G. Ksiazek, S. T. Nichol, High-throughput molecular detection of hemorrhagic fever virus threats with applications for outbreak settings. *J. Infect. Dis.* **196** (suppl. 2), S205–S212 (2007). Medline doi:10.1086/520601
8. M. Panning, T. Laue, S. Olschlager, M. Eickmann, S. Becker, S. Raith, M. C. Courbot, M. Nilsson, R. Gopal, A. Lundkvist, A. Caro, D. Brown, H. Meyer, G. Lloyd, B. M. Kummerer, S. Gunther, C. Drosten, Diagnostic reverse-transcription polymerase chain reaction kit for filoviruses based on the strain collections of all European biosafety level 4 laboratories. *J. Infect. Dis.* **196** (suppl. 2), S199–S204 (2007). Medline doi:10.1086/520600
9. C. M. Malboeuf, X. Yang, P. Charlebois, J. Qu, A. M. Berlin, M. Casali, K. N. Pesko, C. L. Boutwell, J. P. DeVincenzo, G. D. Ebel, T. M. Allen, M. C. Zody, M. R. Henn, J. Z. Levin, Complete viral RNA genome sequencing of ultra-low copy samples by sequence-independent amplification. *Nucleic Acids Res.* **41**, e13 (2013). Medline doi:10.1093/nar/gks794
10. A. Sanchez, S. G. Trappier, B. W. Mahy, C. J. Peters, S. T. Nichol, The virion glycoproteins of Ebola viruses are encoded in two reading frames and are expressed through transcriptional editing. *Proc. Natl. Acad. Sci. U.S.A.* **93**,

3602–3607 (1996).
11. V. E. Volchkov, S. Becker, V. A. Volchkova, V. A. Ternovoj, A. N. Kotov, S. V. Netesov, H. D. Klenk, GP mRNA of Ebola virus is edited by the Ebola virus polymerase and by T7 and vaccinia virus polymerases. *Virology* **214**, 421–430 (1995). Medline doi:10.1006/viro.1995.0052
12. V. A. Volchkova, O. Dolnik, M. J. Martinez, O. Reynard, V. E. Volchkov, Genomic RNA editing and its impact on Ebola virus adaptation during serial passages in cell culture and infection of guinea pigs. *J. Infect. Dis.* **204** (suppl. 3), S941–S946 (2011). Medline doi:10.1093/infdis/jir321
13. G. Dudas, A. Rambaut, Phylogenetic analysis of Guinea 2014 EBOV Ebolavirus outbreak. *PLOS Curr. Outbreaks* **6**, 10.1371/currents.outbreaks.84eefe5ce43ec9dc0bf0670f7b8b417d (2014). doi:10.1371/currents.outbreaks.84eefe5ce43ec9dc0bf0670f7b8b417d
14. J. Kuhn, C. H. Calisher, Eds., *Filoviruses: A Compendium of 40 Years of Epidemiological, Clinical, and Laboratory Studies* (Springer, New York, 2008).
15. M. J. Schreiber, E. C. Holmes, S. H. Ong, H. S. Soh, W. Liu, L. Tanner, P. P. Aw, H. C. Tan, L. C. Ng, Y. S. Leo, J. G. Low, A. Ong, E. E. Ooi, S. G. Vasudevan, M. L. Hibberd, Genomic epidemiology of a dengue virus epidemic in urban Singapore. *J. Virol.* **83**, 4163–4173 (2009). Medline doi:10.1128/JVI.02445-08
16. J. O. Wertheim, S. L. Kosakovsky Pond, Purifying selection can obscure the ancient age of viral lineages. *Mol. Biol. Evol.* **28**, 3355–3365 (2011). Medline doi:10.1093/molbev/msr170
17. S. Y. Ho, M. J. Phillips, A. Cooper, A. J. Drummond, Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol. Biol. Evol.* **22**, 1561–1568 (2005). Medline doi:10.1093/molbev/msi145
18. E. C. Holmes, Patterns of intra- and interhost nonsynonymous variation reveal strong purifying selection in dengue virus. *J. Virol.* **77**, 11296–11298 (2003). Medline doi:10.1128/JVI.77.20.11296-11298.2003
19. J. R. Kugelman, M. S. Lee, C. A. Rossi, S. E. McCarthy, S. R. Radoshitzky, J. M. Dye, L. E. Hensley, A. Honko, J. H. Kuhn, P. B. Jahrling, T. K. Warren, C. A. Whitehouse, S. Bavari, G. Palacios, Ebola virus genome plasticity as a marker of its passaging history: A comparison of in vitro passaging to non-human primate infection. *PLOS ONE* **7**, e50316 (2012). Medline doi:10.1371/journal.pone.0050316
20. S. Günther, M. Asper, C. Röser, L. K. Luna, C. Drosten, B. Becker-Ziaja, P. Borowski, H. M. Chen, R. S. Hosmane, Application of real-time PCR for testing antiviral compounds against Lassa virus, SARS coronavirus and Ebola virus in vitro. *Antiviral Res.* **63**, 209–215 (2004). Medline doi:10.1016/j.antiviral.2004.05.001
21. G. Grard, R. Biek, J. J. Muyembe Tamfum, J. Fair, N. Wolfe, P. Formenty, J. Paweska, E. Leroy, Emergence of divergent Zaire ebola virus strains in Democratic Republic of the Congo in 2007 and 2008. *J. Infect. Dis.* **204** (suppl. 3), S776–S784 (2011). Medline doi:10.1093/infdis/jir364
22. G. P. Kobinger, A. Leung, J. Neufeld, J. S. Richardson, D. Falzarano, G. Smith, K. Tierney, A. Patel, H. M. Weingartl, Replication, pathogenicity, shedding, and transmission of Zaire ebolavirus in pigs. *J. Infect. Dis.* **204**, 200–208 (2011). Medline doi:10.1093/infdis/jir077
23. T. Hoenen, S. Jung, A. Herwig, A. Groseth, S. Becker, Both matrix proteins of Ebola virus contribute to the regulation of viral genome replication and transcription. *Virology* **403**, 56–66 (2010). doi:10.1016/j.virol.2010.04.002
24. J. A. Blow, C. N. Mores, J. Dyer, D. J. Dohm, Viral nucleic acid stabilization by RNA extraction reagent. *J. Virol. Methods* **150**, 41–44 (2008). Medline doi:10.1016/j.jviromet.2008.02.003
25. A. R. Trombley, L. Wachter, J. Garrison, V. A. Buckley-Beason, J. Jahrling, L. E. Hensley, R. J. Schoepp, D. A. Norwood, A. Goba, J. N. Fair, D. A. Kulesh, Comprehensive panel of real-time TaqMan polymerase chain reaction assays for detection and absolute quantification of filoviruses, arenaviruses, and New World hantaviruses. *Am. J. Trop. Med. Hyg.* **82**, 954–960 (2010). Medline doi:10.4269/ajtmh.2010.09-0636
26. J. D. Morlan, K. Qu, D. V. Sinicropi, Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. *PLOS ONE* **7**, e42882 (2012). Medline doi:10.1371/journal.pone.0042882
27. X. Adiconis, D. Borges-Rivera, R. Satija, D. S. DeLuca, M. A. Busby, A. M. Berlin, A. Sivachenko, D. A. Thompson, A. Wysoker, T. Fennell, A. Gnirke, N. Pochet, A. Regev, J. Z. Levin, Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat. Methods* **10**, 623–629 (2013). Medline doi:10.1038/nmeth.2483
28. L. Jiang, F. Schlesinger, C. A. Davis, Y. Zhang, R. Li, M. Salit, T. R. Gingeras, B. Oliver, Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* **21**, 1543–1551 (2011). Medline doi:10.1101/gr.121095.111
29. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004). Medline doi:10.1093/nar/gkh340
30. P. Cingolani, A. Platts, L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, D. M. Ruden, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012). Medline doi:10.4161/fly.19695
31. A. Stamatakis, T. Ludwig, H. Meier, RAxML-III: A fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21**, 456–463 (2005). doi:10.1093/bioinformatics/bti191
32. F. Ronquist, J. Huelsenbeck, MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003). doi:10.1093/bioinformatics/btg180
33. A. J. Drummond, M. A. Suchard, D. Xie, A. Rambaut, Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012). Medline doi:10.1093/molbev/mss075
34. M. Hasegawa, H. Kishino, T. Yano, Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985). Medline doi:10.1007/BF02101694
35. Z. Yang, Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J. Mol. Evol.* **39**, 306–314 (1994). Medline doi:10.1007/BF00160154
36. M. S. Gill, P. Lemey, N. R. Faria, A. Rambaut, B. Shapiro, M. A. Suchard, Improving Bayesian population dynamics inference: A coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30**, 713–724 (2013). Medline doi:10.1093/molbev/mss265
37. A. J. Drummond, S. Y. Ho, M. J. Phillips, A. Rambaut, Relaxed phylogenetics and dating with confidence. *PLOS Biol.* **4**, e88 (2006). Medline doi:10.1371/journal.pbio.0040088
38. G. Baele, P. Lemey, S. Vansteelandt, Make the most of your samples: Bayes factor estimators for high-dimensional models of sequence evolution. *BMC Bioinform.* **14**, 85 (2013). doi:10.1186/1471-2105-14-85
39. M. A. Ferreira, M. C. O'Donovan, Y. A. Meng, I. R. Jones, D. M. Ruderfer, L. Jones, J. Fan, G. Kirov, R. H. Perlis, E. K. Green, J. W. Smoller, D. Grozeva, J. Stone, I. Nikolov, K. Chambert, M. L. Hamshere, V. L. Nimgaonkar, V. Moskvina, M. E. Thase, S. Caesar, G. S. Sachs, J. Franklin, K. Gordon-Smith, K. G. Ardlie, S. B. Gabriel, C. Fraser, B. Blumenstiel, M. Defelice, G. Breen, M. Gill, D. W. Morris, A. Elkin, W. J. Muir, K. A. McGhee, R. Williamson, D. J. MacIntyre, A. W. MacLean, C. D. St, M. Robinson, M. Van Beck, A. C. Pereira, R. Kandaswamy, A. McQuillin, D. A. Collier, N. J. Bass, A. H. Young, J. Lawrence, I. N. Ferrier, A. Anjorin, A. Farmer, D. Curtis, E. M. Scolnick, P. McGuffin, M. J. Daly, A. P. Corvin, P. A. Holmans, D. H. Blackwood, H. M. Gurling, M. J. Owen, S. M. Purcell, P. Sklar, N. Craddock, Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat. Genet.* **40**, 1056–1058 (2008). Medline doi:10.1038/ng.209
40. M. Mehedi, D. Falzarano, J. Seebach, X. Hu, M. S. Carpenter, H. J. Schnittler, H. Feldmann, A new Ebola virus nonstructural glycoprotein expressed through RNA editing. *J. Virol.* **85**, 5406–5414 (2011). Medline doi:10.1128/JVI.02190-10
41. T. R. Gibb, D. A. Norwood Jr., N. Woollen, E. A. Henchal, Development and evaluation of a fluorogenic 5′ nuclease assay to detect and differentiate between Ebola virus subtypes Zaire and Sudan. *J. Clin. Microbiol.* **39**, 4125–4130 (2001). Medline doi:10.1128/JCM.39.11.4125-4130.2001
42. J. M. Morvan, V. Deubel, P. Gounon, E. Nakouné, P. Barrière, S. Murri, O. Perpète, B. Selekon, D. Coudrier, A. Gautier-Hion, M. Colyn, V. Volehkov, Identification of Ebola virus sequences present as RNA or DNA in organs of terrestrial small mammals of the Central African Republic. *Microbes Infect.* **1**, 1193–1201 (1999). Medline doi:10.1016/S1286-4579(99)00242-7
43. A. Sanchez, T. G. Ksiazek, P. E. Rollin, M. E. Miranda, S. G. Trappier, A. S. Khan, C. J. Peters, S. T. Nichol, Detection and molecular characterization of Ebola viruses causing disease in human and nonhuman primates. *J. Infect. Dis.* **179** (suppl. 1), S164–S169 (1999). Medline doi:10.1086/514282
44. M. Weidmann, E. Mühlberger, F. T. Hufert, Rapid detection protocol for

filoviruses. *J. Clin. Virol.* **30**, 94–99 (2004). Medline doi:10.1016/j.jcv.2003.09.004

Acknowledgments: We thank the Sierra Leone MoHS (Hon. Minister M. Kargbo, B. Kargbo, M.A. Vandi, A. Jambai), the Kenema District Health Management Team and Lassa fever program for their efforts in outbreak response. We thank P. Cingolani, Y.-C. Wu, M. Lipsitch, S. Günther, S. Baize, N. Wauquier, J. Bangura, V. Lungay, L. Hensley, J. Johnson, M. Voorhees, A. O'Hearn, and R. Schoepp, L. Gaffney, J. Kuhn, S.C. Sealfon, J.B. Shapiro, C. Edwards, Sabeti lab members for technical support and feedback. This project has been funded in part by NIH 1DP2OD006514-01 and NIAID HHSN272200900049C. RS is supported by NSF GRFP, SW by NIH GM080177; CH by NIH 1U01HG007480-01 and the World Bank; AR by EU FP7/2007-2013 278433-PREDEMICS and ERC 260864; and GD by NERC D76739X. Sequence data are available at NCBI (NCBI BioGroup: PRJNA257197). Sharing of RNA samples used in this study requires approval from the Sierra Leone Ministry of Health and Sanitation.

Supplementary Materials
www.sciencemag.org/cgi/content/full/science.1259657/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S11
Tables S1 to S4
Supplementary files S1 to S4
References (*19*–*44*)

5 August 2014; accepted 21 August 2014

Published online 28 August 2014
10.1126/science.1259657

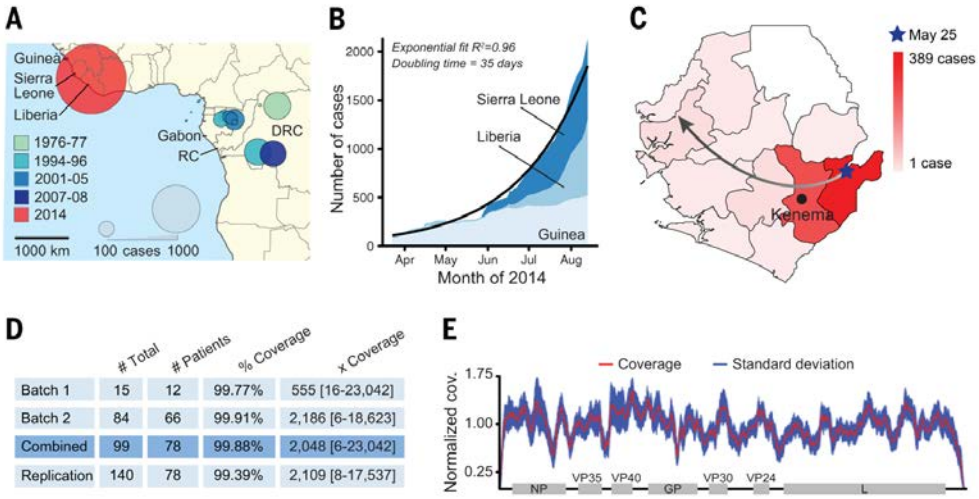

Fig. 1. Ebola outbreaks, historical and current. (**A**) Historical EVD outbreaks, colored by decade. Circle area represents total number of cases (RC = Republic of the Congo; DRC = Democratic Republic of Congo). (**B**) 2014 outbreak growth (confirmed, probable and suspected cases). (**C**) Spread of EVD in Sierra Leone by district. The gradient denotes number of cases and the arrows depict likely direction. (**D**) EBOV samples from 78 patients were sequenced in two batches, totaling 99 viral genomes (Replication = technical replicates (*6*)). Mean coverage and median depth of coverage with range are shown. (**E**) Combined normalized (to the sample average) coverage across sequenced EBOV genomes.

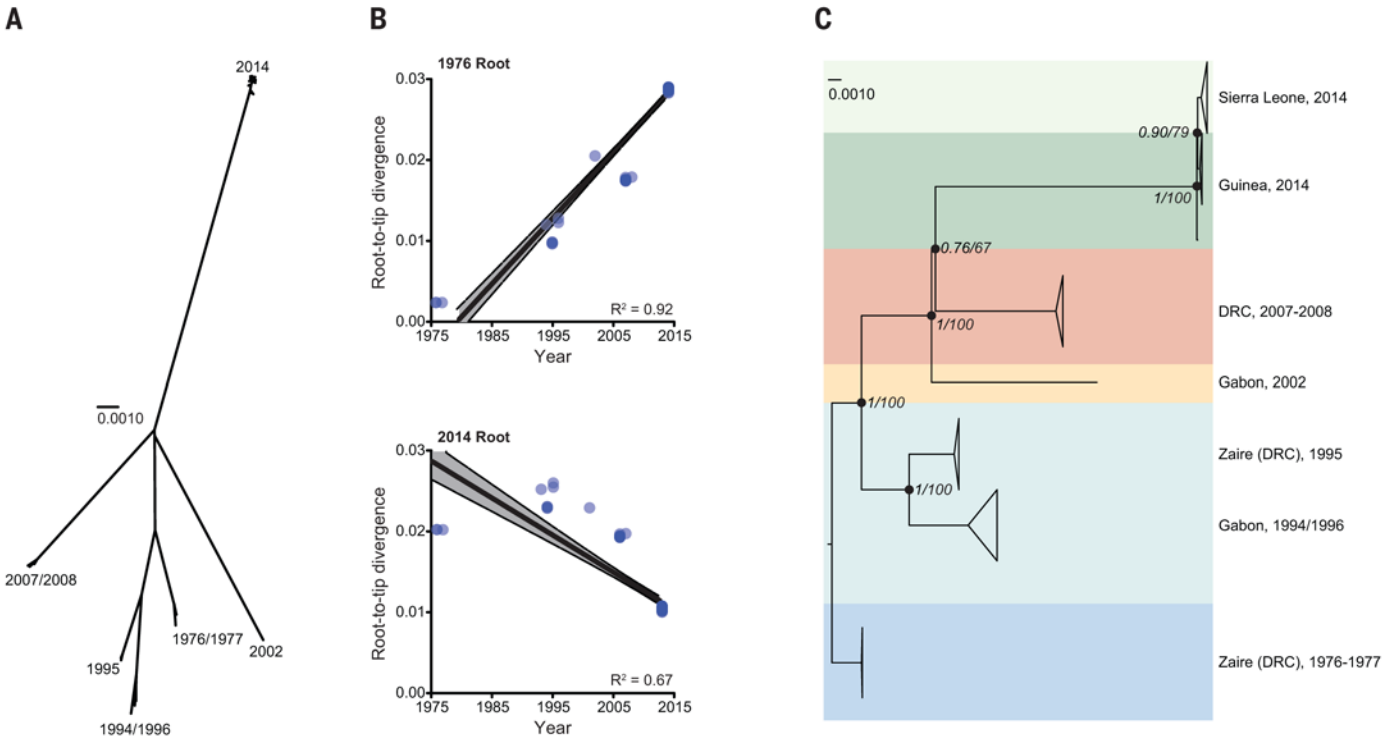

Fig. 2. Relationship between outbreaks. (**A**) Unrooted phylogenetic tree of EBOV samples; each major clade corresponds to a distinct outbreak (scale bar = nucleotide substitutions/site). (**B**) Root-to-tip distance correlates better with sample date when rooting on the 1976 branch (R^2 = 0.92, top) than on the 2014 branch (R^2 = 0.67, bottom). (**C**) Temporally rooted tree from (A).

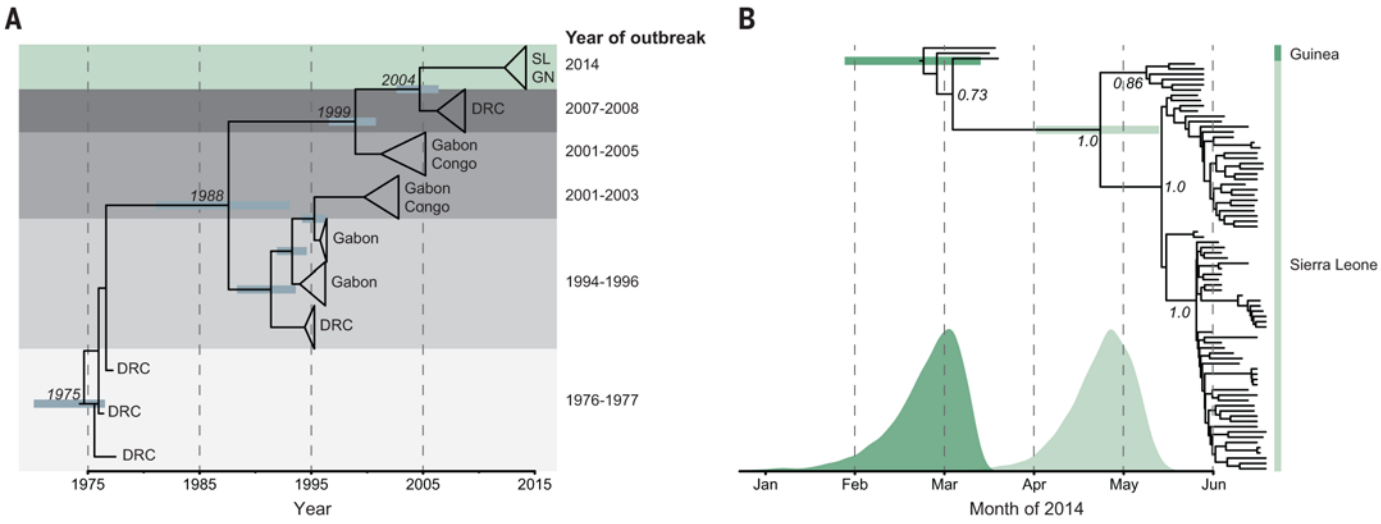

Fig. 3. Molecular dating of the 2014 outbreak. (**A**) BEAST dating of the separation of the 2014 lineage from Middle African lineages (SL = Sierra Leone; GN = Guinea; DRC = Democratic Republic of Congo; tMRCA: Sep 2004, 95% HPD: Oct 2002 - May 2006). (**B**) BEAST dating of the tMRCA of the 2014 West African outbreak (tMRCA: Feb 23, 95% HPD: Jan 27 - Mar 14) and the tMRCA of the Sierra Leone lineages (tMRCA: Apr 23, 95% HPD: Apr 2 - May 13); probability distributions for both 2014 divergence events overlayed below. Posterior support for major nodes is shown.

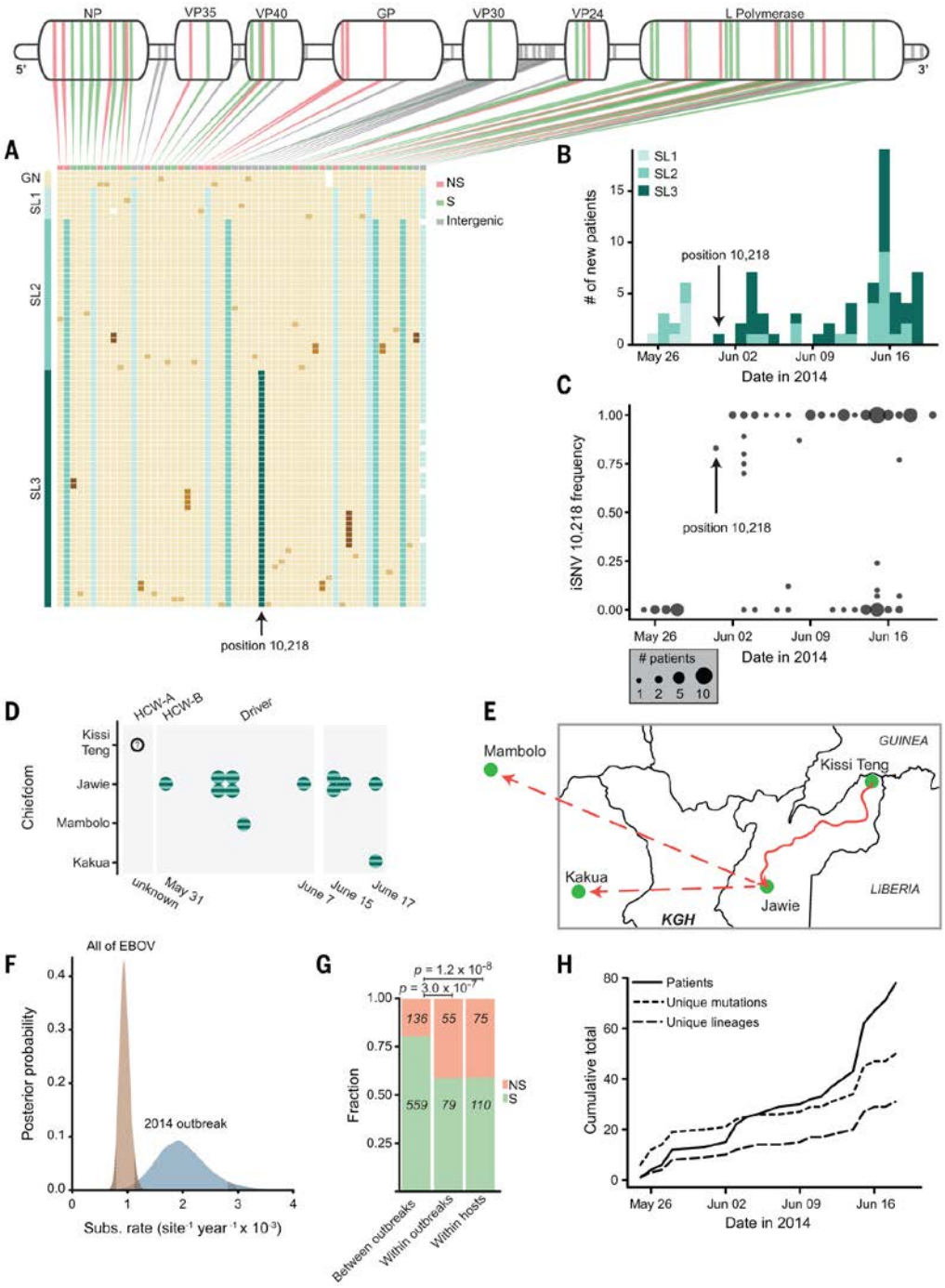

Fig. 4. Viral dynamics during the 2014 outbreak. (**A**) Mutations, one patient sample per row; beige = identical to Kissidougou Guinean sequence (accession KJ660346). The top row shows the type of mutation (green: synonymous, pink: nonsynonymous, intergenic: gray), with genomic locations indicated above. Clusters assignments are shown on left. (**B**) Number of EVD-confirmed patients per day, colored by cluster (arrow: first appearance of the derived allele at position 10,218, distinguishing clusters 2 and 3). (**C**) Intrahost frequency of SNP 10,218 in all 78 patients (absent in 28 patients, polymorphic in 12, fixed in 38). (**D** and **E**) 12 patients carrying iSNV 10,218 cluster geographically and temporally (HCW-A = unsequenced health care worker, Driver drove HCW-A from Kissi Teng to Jawie, then continued alone to Mambolo, HCW-B treated HCW-A). (**F**) Substitution rates within the 2014 outbreak and between all EVD outbreaks. (**G**) Proportion of nonsynonymous changes observed on different time scales (green = synonymous; pink = nonsynonymous). (**H**) Acquisition of genetic variation over time. 50 mutational events (short dashes) and 29 new viral lineages (long dashes) were observed (intrahost variants not included).