



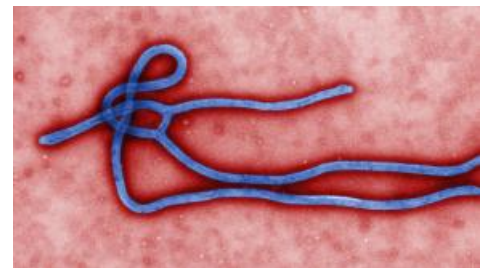
# Ebola outbreak traced to the funeral of traditional healer

## EDITOR'S INTRODUCTION

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

annotated by Hazel O'Connor

Ebola virus disease (EVD) is a severe and most often fatal disease in humans, with an average fatality rate of 78%. The recent (2014) outbreak in West Africa is the largest outbreak to date, with 1229 people reported to have died at the time of publication (August 2014). Since then, this number has risen to more than 10,000 deaths. In the largest study of its kind, the authors explored the origins of the West African outbreak by sequencing viral genomes from patients in Sierra Leone. It was concluded that two genetically distinct strains of viruses were introduced from Guinea to Sierra Leone in late April, with the virus continuing to spread by human-to-human transmission. This study was instrumental to the understanding of the outbreak and led to heightened efforts to contain the outbreak. Tragically, five co-authors died of EVD before the paper was published.



## PAPER DETAILS

**Original title:** Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak

**Authors:** Pardis Sabeti

**Original publication date:** 09/12/2014

**Reference:** Vol. 345 no. 6202 pp. 1369-1372

**Issue name:** Science

**DOI:** 10.1126/science.1259657

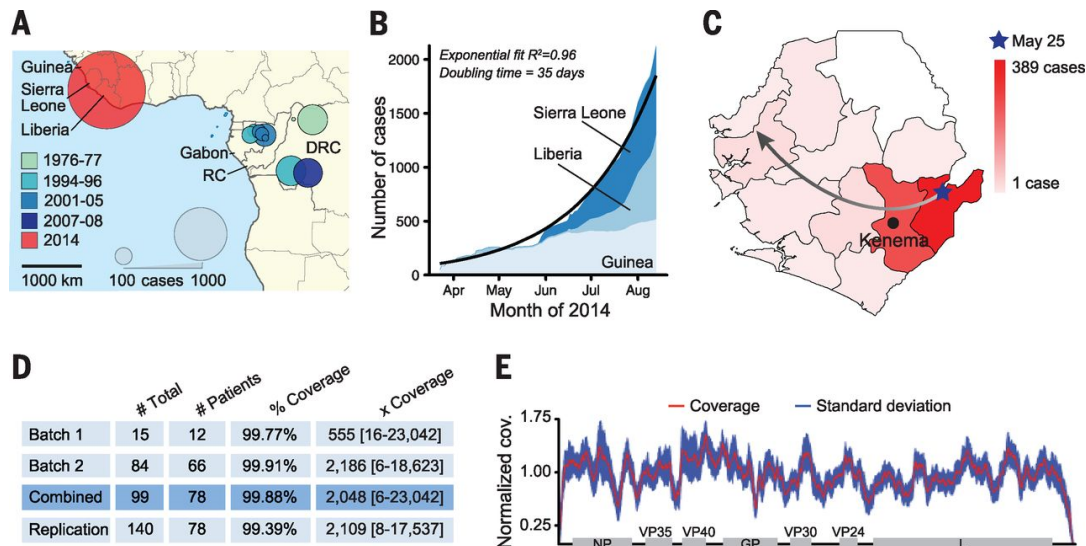
## ABSTRACT

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 ~2000× 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 central African lineages around 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. Because 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.

## REPORT

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 central 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 19 August 2014, 2240 cases and 1229 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.



**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; the arrow depicts 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 coverage (normalized to the sample average) across sequenced EBOV genomes.

### Panel A

The scale of the current outbreak is illustrated.

The size of the dots represents the size of the outbreaks (number of cases).

Note the difference in the size of the outbreak and also the geographical location (Western Africa) compared with previous outbreaks (Central Africa).

## **Panels B and C**

The outbreak is growing (B) and spreading (C) quickly.

## **Panel D: Author's experiments**

This outbreak appeared to be remarkably different from previous outbreaks (A-C). Patient samples were sequenced to gain more information about the virus.

Samples from 78 patients were sequenced; however, samples from some patients were sequenced more than once (# total), which gave the authors more information on the evolution of the virus in a single individual over time. Some samples were sequenced more than once as experimental controls (replication = technical replicates).

A technique called deep sequencing was used to generate the sequencing results of the EBOV genome. X coverage represents the depth of coverage; that is the total number of times a given nucleotide has been sequenced by independent reads. For example, batch 1 achieved 555 X coverage (average coverage across the entire genome), with a range of 16 to 23,042. (The minimum times a region was sequenced was 16 times, the maximum was 23,042.)

Percentage coverage is an estimation of the amount of the genome that was successfully covered by the sequencing reads. The depth and percentage coverage achieved here are likely to allow the authors to detect any mutations that are present.

## **Panel E**

Ebola (EBOV) is a single stranded RNA virus that consists of 7 structural proteins (NP, VP35, VP40, GP, VP30, VP24, L).

Depicted here is the sequence coverage achieved for all samples across the entire genome.

During the sequencing process, certain parts of the genome are more difficult to sequence than others, therefore they may be underrepresented in the final results compared with others; normalized cov. (coverage) accounts for the over- or underrepresentation of nucleotides compared with others.

The standard deviation is a measure of the variation within an individual sample, compared with the group as a whole.

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 polymerase chain reaction (PCR)–based EBOV diagnostics (8) (fig. S2); all tests were negative through early May. On 25 May, 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 from 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 United States according to 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).

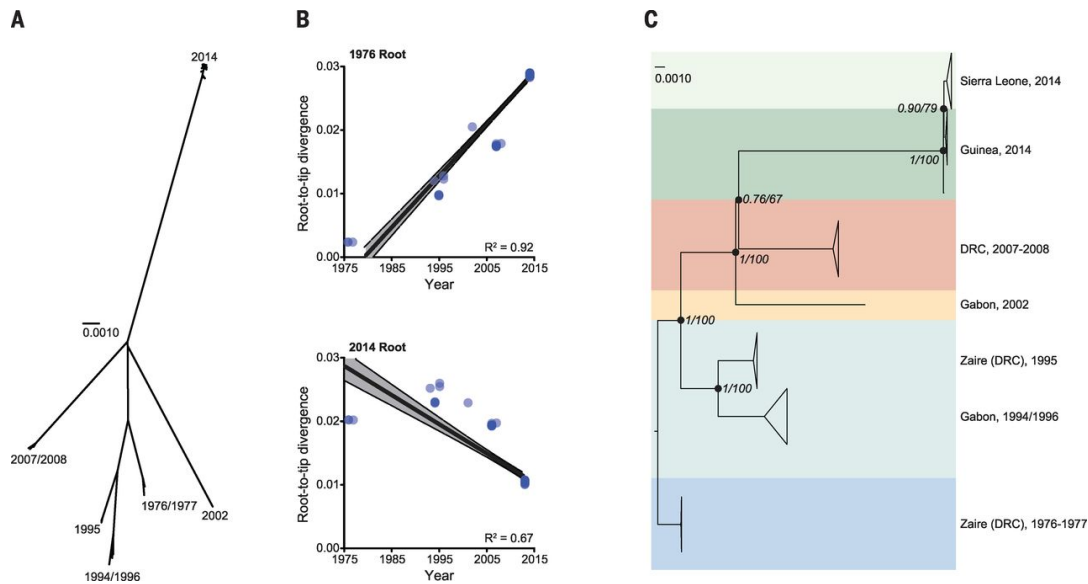
In total, we generated 99 EBOV genome sequences from 78 confirmed EVD patients, representing more than 70% of the EVD patients diagnosed in Sierra Leone from late May to mid-June; we used multiple extraction methods or time points for 13 patients (table S2). Median coverage was >2000×, spanning more than 99.9% of EBOV coding regions (Fig. 1, D and E, and table S2).

We combined the 78 Sierra Leonean sequences with three published Guinean samples (3) [correcting 21 likely sequencing errors in the latter (6)] to obtain a data set of 81 sequences. They reveal 341 fixed substitutions (35 nonsynonymous, 173 synonymous, and 133 noncoding) between the 2014 EBOV and all previously published EBOV sequences, with an additional 55 single-nucleotide polymorphisms (SNPs; 15 nonsynonymous, 25 synonymous, and 15 noncoding), fixed within individual patients, within the West African outbreak. Notably, the Sierra Leonean genomes differ from PCR probes for four 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 characterized in patients (6).

Phylogenetic comparison to all 20 genomes from earlier outbreaks suggests that the 2014 West African virus likely spread from central Africa within the past decade. Rooting the phylogeny using divergence from 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  $8 \times 10^{-4}$  per site per 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, around 2004 (Fig. 2C and Fig. 3A), which

supports the hypothesis that each outbreak represents an independent zoonotic event from the same genetically diverse viral population in its natural reservoir.



**Fig. 2. Relationship between outbreaks. (A) Unrooted phylogenetic tree of EBOV samples; each major clade corresponds to a distinct outbreak (scale bar = nucleotide substitutions per 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).**

### Question

The key questions that authors wished to address were:

i) Where did the virus come from?

ii) How is it evolving?

#### Panel A

An unrooted tree specifies relationships among species, but not from an ancestor; this type of tree does not assume a known ancestor.

The branches represent species being compared, and the internal nodes (where the branches meet) are ancestral strains (where the species are estimated to diverged from each other).

Branch length describes the evolutionary divergence between two nodes (genetic change).

In Figure 2A, the length of the branches represents the number of nucleotide changes.

The length of the 2014 branch demonstrates that the 2014 EBOV strain is acquiring mutations more rapidly and has therefore diverged more than previous outbreaks.

## **Panel B**

Providing evidence for the accumulation of sequence divergence consists of performing a regression between root-to-tip divergence and sampling date in a "classical" phylogenetic tree (with a strict molecular clock assumption).

The R-squared of the regression indicates the amount of molecular clock signal and is referred to as the "root-to-tip" method. Estimating the root-to-tip divergence provides us the R-squared of the regression between root-to-tip divergence.

The R-squared indicates the amount of sequence divergence explained by the sampling date.

The authors generated phylogenies using a ML likelihood approach (using the software PhyML v.3.0) and then estimated the molecular clock by performing a regression between root-to-tip distance in the ML tree and the date of sampling of each sequence using the software Path-O-Gen v1.3.

Trees were rooted at the position that was likely to be the most compatible with the assumption of the molecular clock.

This method estimated the amount of variation in genetic distances that can be explained by the sampling time.

This figure tells us that the best approach for studying the evolution of the EBOV is to assume that the oldest ancestor is the 1976 species.

## **Panel C**

This temporally rooted tree assumes that the oldest ancestor is the Zaire 1976 strain and tracks changes over time.

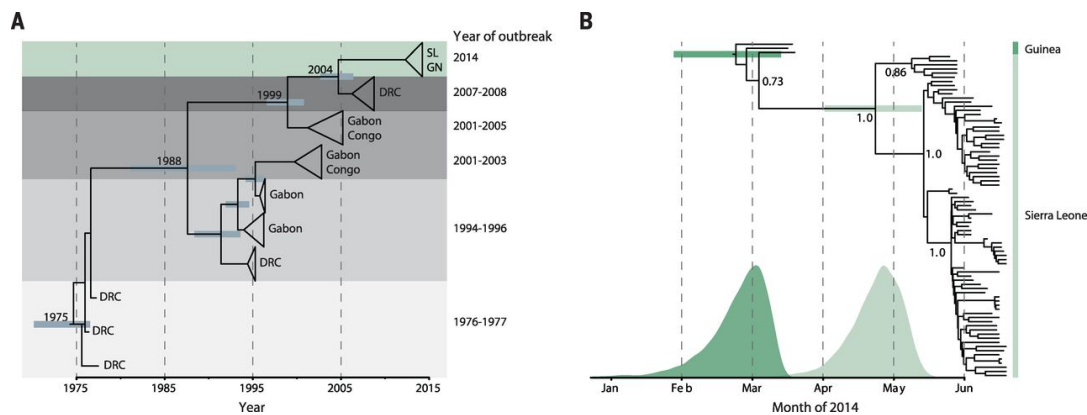
The nodes of the tree are the theoretical ancestors, an estimate of the time at which the viruses diverged, assuming a constant rate of mutation.

The best fit model in fig 2B is that the Zaire 1976 species is the oldest ancestor and therefore, the root of the tree.

The line segment with the number "0.0010" shows the length of branch that represents an amount genetic change of 0.0010. The units of branch length are usually nucleotide substitutions per site— that is the number of changes or "substitutions" divided by the length of the sequence.

Fig 2C suggests that the 2014 outbreak is most closely related to the DRC (Democratic Republic of Congo) 2007-2008 outbreak strains.

The Guinea and Sierra Leone 2014 strains are very closely related, compared with other outbreaks, providing evidence that the Sierra Leone outbreak came from Guinea and that human-to-human transmission is occurring, rather than the virus being transmitted from the reservoir.



**Fig. 3. Molecular dating of the 2014 outbreak. (A) BEAST dating of the separation of the 2014 lineage from central African lineages [SL, Sierra Leone; GN, Guinea; DRC, Democratic Republic of Congo; time of most recent common ancestor (tMRCA), September 2004; 95% highest posterior density (HPD), October 2002 to May 2006]. (B) BEAST dating of the tMRCA of the 2014 West African outbreak (23 February; 95% HPD, 27 January to 14 March) and the tMRCA of the Sierra Leone lineages (23 April; 95% HPD, 2 April to 13 May). Probability distributions for both 2014 divergence events are overlaid below. Posterior support for major nodes is shown.**

### Molecular dating

Molecular dating is a technique that classifies the evolutionary relationships between species with relation to time, in order to estimate i) how long species have existed and ii) how long ago certain species diverged from each other. The simplest model, the molecular clock, assumes that the rate of substitution remains constant over time.

### BEAST dating

Several models, such as BEAST dating, have been developed and implemented for inferring divergence times without assuming a strict molecular clock. This is favorable, because many factors can influence the rate of substitution such as mutation rate, population size, and selection.

The authors obtained the phylogenetic trees using Bayesian inference methods, a statistical approach for assessing the probability of a hypothesis given the evidence at hand. In order to obtain reliable estimates of speciation times, the time of sampling for the EBOV phylogenies was incorporated into the program BEAST v1.8. This was coupled with Markov chain Monte Carlo (MCMC), for approximating the posterior probability distribution of parameters.

In statistical terms, the posterior probability is the probability of event A occurring given that event B has occurred.

### Panel A

Through knowledge of the time of the diagnostic test and of the sequenced virus associated with that outbreak, it is estimated that the DRC 2007-2008 and SL, GN 2014 outbreaks had a common ancestor in 2004. The triangular symbols represent the sample sizes available for each particular outbreak.

### Panel B

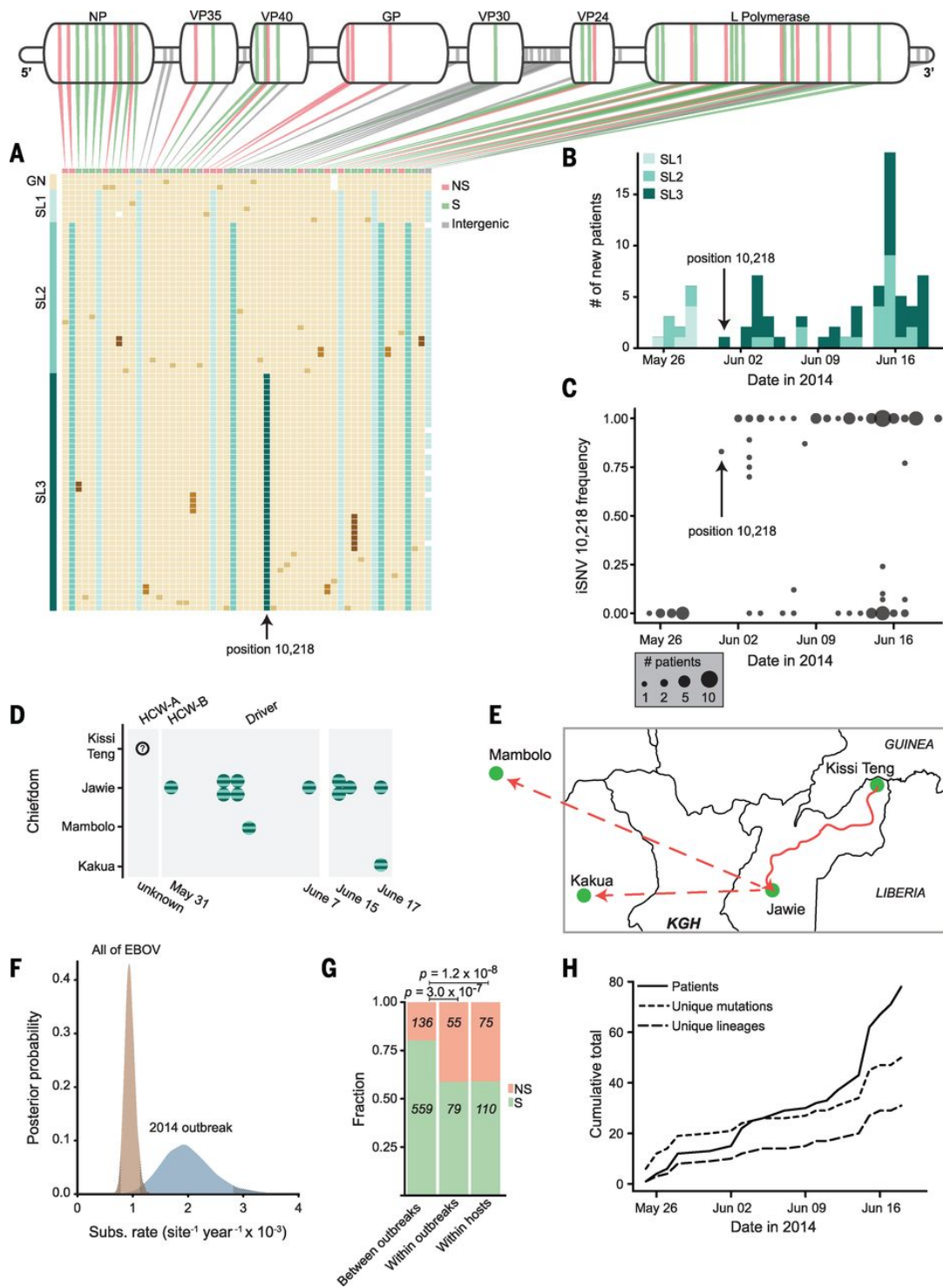
In Figure 3B, the authors take a closer look at the molecular clock of the most recent outbreak.

Using BEAST dating to correlate the timing with sequence comparisons with statistical methods, it is believed that the West African outbreak began on February 23rd (Guinea; dark green color) and the Sierra Leone variant (light green color) of the virus emerged on 23 April 2014. Ninety-five percent HPD refers to a 95% confidence in the accuracy of these dates.

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), 3 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 that 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), predating their co-appearance in Sierra Leone in late May (Fig. 4B); this finding suggests that 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.





**Fig. 4. Viral dynamics during the 2014 outbreak. (A)** Mutations, one patient sample per row; beige blocks indicate identity with the Kissidougou Guinean sequence (GenBank accession KJ660346). The top row shows the type of mutation (green, synonymous; pink, nonsynonymous; gray, intergenic), with genomic locations indicated above. Cluster assignments are shown at the left. **(B)** Number of EVD-confirmed patients per day, colored by cluster. Arrow indicates the first appearance of the derived allele at position 10,218, distinguishing clusters 2 and 3. **(C)** Intra-host frequency of SNP 10,218 in all 78 patients (absent in 28 patients, polymorphic in 12, fixed in 38). **(D and E)** Twelve 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). KGH = location of Kenema Government Hospital. (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. Fifty mutational events (short dashes) and 29 new viral lineages (long dashes) were observed (intra-host variants not included).

#### **Panel A**

The sequencing data is probed for variation in the viral genome.

Sequencing of the EBOV genome documented changes in both the protein-coding regions and the intergenic regions of the genome.

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 (SL1 and SL2).

All subsequent diversity in Sierra Leone accumulated on the background of those two lineages (Fig. 4A), consistent with epidemiological information from tracing contacts.

A new lineage of EBOV emerged (SL3), characterized by variation at position 10,218.

#### **Panel B**

The coappearance in Sierra Leone in late May (Fig. 4B) of the two lineages.

In Figure 3B, molecular dating placed the divergence of SL1 and SL2 lineages in late April, predating their coappearance in Sierra Leone in late May.

This led the authors to postulate that the funeral attendees were most likely infected by two lineages then circulating in Guinea.

In this figure, it is observed that by June 16, the newest lineage SL3, has become the predominant lineage.

#### **Panel C**

The frequency of variation at position 10,218 increased significantly in late May, and became dominant in early June.

#### **Panels D and E**

Gathering epidemiological information was imperative to tracking the origin and spread of EBOV lineages of the 2014 outbreak.

By tracking the movements of the 12 patients who are believed to have spread EBOV to Sierra Leone (Mambolo and Kakua), the authors were able to further support the conclusions of their sequencing data.

#### **Panels F, G, and H**

By comparing the number of changes (substitutions) among sequenced samples in the current outbreak to previous outbreaks, it appears that substitution rate for the current outbreak is twice as high.

Upon further analysis of these substitutions, it is found that changes are more frequently nonsynonymous during this outbreak, compared to all outbreaks (Fig. 4G).

These nonsynonymous mutations could potentially indicate selective pressure, but there are amount of mutations are not so extensive that other explanations may be ruled out.

The unprecedented rate of acquisition of nonsynonymous mutations and the emergence of a new lineage in such a short space of time is a cause for concern however (4H), and may suggest that continued progression of this epidemic could afford an opportunity for viral adaptation.

#### **What's left to learn about the Ebola outbreak?**

The authors released all of the genetic information gained in this article to the public, to enable fellow scientists to gain further insights as to how the outbreak can be contained. Can a successful cure and vaccines be developed?

It remains to be seen if and/or how the virus is evolving to become a more deadly pathogen.

Combining sequence data with more recent patient samples may shed more light on the evolution of Ebola.

Finding the reservoir is a major challenge of the field. Is it really bats?

Patterns in observed intrahost and interhost variation provide important insight 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 12 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 support 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, provides a starting point for such studies (table S4).

To aid in relief efforts and facilitate rapid global research, we have immediately released all sequence data as it is 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 and lost their battle with the disease before this manuscript could be published: Mohamed Fullah, Mbalu Fonnio, Alex Moigboi, Alice Kovoma, and S. Humarr Khan. We wish to honor their memory.

## Supplementary Materials

[www.sciencemag.org/content/345/6202/1369/suppl/DC1](http://www.sciencemag.org/content/345/6202/1369/suppl/DC1)

## Materials and Methods

### Supplementary Text

Figs. S1 to S11

Tables S1 to S4

Supplementary files S1 to S4

References (19–44)

## 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).
2. J. Burke, Ebola haemorrhagic fever in Zaire, 1976. *Bull. World Health Organ.* **56**, 271–293 (1978).
3. S. Baize *et al.*, Emergence of Zaire Ebola virus disease in Guinea—Preliminary report. *N. Engl. J. Med.* **10.1056/NEJMoa1404505** (2014).
4. WHO, (2014), [www.who.int/csr/don/archive/disease/ebola/en/](http://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).
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).
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).
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.**

10. 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).
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).
13. G. Dudas, A. Rambaut, Phylogenetic analysis of Guinea 2014 EBOV Ebolavirus outbreak. *PLOS Curr. Outbreaks* **6**, 10.1371/currents.outbreaks.84eefe5ce43ec9dc0bf0670f7b8b417d (2014).
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).
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).
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).
18. E. C. Holmes, Patterns of intra- and interhost nonsynonymous variation reveal strong purifying selection in dengue virus. *J. Virol.* **77**, 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).
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).
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).
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).

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).
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).
25. 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).
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).
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).
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).
29. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
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).
31. Stamatakis, T. Ludwig, H. Meier, RAxML-III: A fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21**, 456–463 (2005).
32. F. Ronquist, J. Huelsenbeck, MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
33. 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).
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).
35. Z. Yang, Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J. Mol. Evol.* **39**, 306–314 (1994).
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).
37. J. Drummond, S. Y. Ho, M. J. Phillips, A. Rambaut, Relaxed phylogenetics and dating with confidence. *PLOS Biol.* **4**, e88 (2006).

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).
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).
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).
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).
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).
43. 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).
44. M. Weidmann, E. Mühlberger, F. T. Hufert, Rapid detection protocol for filoviruses. *J. Clin. Virol.* **30**, 94–99 (2004).
45. Acknowledgments: We thank the Office of the President of Sierra Leone (President E. Koroma, M. Jones, S. Blyden), the Sierra Leone Ministry of Health and Sanitation (Minister M. Kargbo, B. Kargbo, M. A. Vandj, A. Jambai), the Kenema District Health Management Team, and the 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, R. Schoepp, L. Gaffney, J. Kuhn, S. C. Sealfon, J. B. Shapiro, C. Edwards, and Sabeti lab members for technical support and feedback. Supported by the NSF Graduate Research Fellowship Program (R.S.G.S.), NIH grant GM080177 (S. Wohl), NIH grant 1U01HG007480-01 and the World Bank (C.H.), European Union grant FP7/2007-2013 278433-PREDEMICS and European Research Council grant 260864 (A.R.), Natural Environment Research Council grant D76739X (G.D.), NIH grant 1DP2OD006514-01, and National Institute of Allergy and Infectious Diseases grant HHSN272200900049C. 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

## TOPICS

[Diseases and disorders](#) [Ecology](#) [Microbiology](#)

## RELATED PAPERS



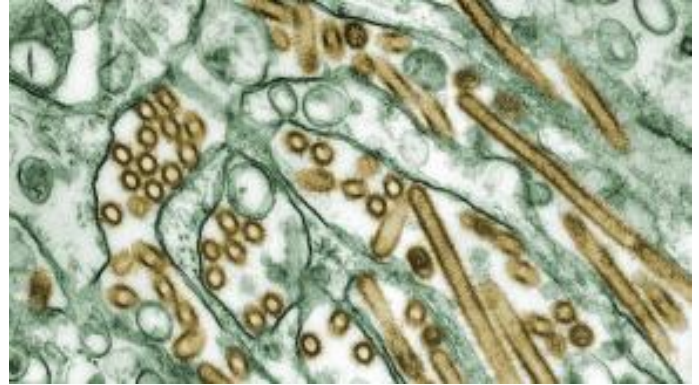
### One vaccine to rule them all

Long-term measles-induced immunomodulation increases overall childhood infectious disease mortality

Annotated by Sara Reynolds

Annotated on 07/21/2016

Originally published 05/08/2015



### Mimicking a bundle of waste: Influenza A virus strategic attack

Influenza A virus uses the aggresome processing machinery for host cell entry

Annotated by Jole Fiorito

Annotated on 01/22/2016

Originally published 10/24/2014



## ABOUT

[About this project](#)

[Partners](#)

[Help](#)

## CONNECT

[Follow us on Twitter](#)





