

## Mapping the clinical outcomes and genetic evolution of Ebola virus in Sierra Leone

Tao Li, Hong-Wu Yao, Di Liu, Hong-Guang Ren, Yi Hu, David Kargbo, Yue Teng, Yong-Qiang Deng, Hui-Jun Lu, Xiong Liu, Kun Liu, Li-Qun Fang, Nian-Zhi Ning, Gary Wong, Foday Daffae, Abdul Kamara, AiPing Wu, Tai-Jiao Jiang, Zhan Li, Jie Huang, Yu Sun, Jun Qian, Brima Kargbo, Jia-Fu Jiang, Hui Wang, Wu-Chun Cao

*JCI Insight*. 2017;2(15):e88333. <https://doi.org/10.1172/jci.insight.88333>.

Research Article

Genetics

Sierra Leone was the most severely affected country in Western Africa during the 2013–2016 outbreak of Ebola virus disease (EVD). Previous genome surveillance studies have revealed the origin, diversity, and evolutionary dynamics of the Ebola virus (EBOV); however, the information regarding EBOV sequences is insufficient, especially the clinical outcomes, given that the correlation between the clinical outcomes and the genetic evolution of EBOV is still not clear. Here, we collected and curated a comprehensive data set that includes 514 EBOV genome sequences from patients with confirmed EVD (including 60 sequences not previously studied), >87.5% of which have residence information and definitive clinical outcomes. Phylogenetic reconstruction revealed 11 lineages of EBOV in Sierra Leone. The median-joining haplotype network showed that haplotypes that are associated with lethal outcomes tend to contribute more to the spread of the EBOV in Sierra Leone than those with live outcomes. Analyses of the spatial-temporal distribution unraveled the lineage-distinctive distribution patterns. Different viral lineages have different case fatality rates (CFRs) during the same stage of the outbreak, implying that several lineages featuring SNPs may correlate with increased/decreased CFRs. This study provides invaluable data sets of EBOV infection and highlights the potential SNPs for further in-depth investigation.

Find the latest version:

<https://jci.me/88333/pdf>



# Mapping the clinical outcomes and genetic evolution of Ebola virus in Sierra Leone

Tao Li,<sup>1</sup> Hong-Wu Yao,<sup>1</sup> Di Liu,<sup>2</sup> Hong-Guang Ren,<sup>1,3</sup> Yi Hu,<sup>1</sup> David Kargbo,<sup>4</sup> Yue Teng,<sup>1</sup> Yong-Qiang Deng,<sup>1</sup> Hui-Jun Lu,<sup>5</sup> Xiong Liu,<sup>1</sup> Kun Liu,<sup>1</sup> Li-Qun Fang,<sup>1</sup> Nian-Zhi Ning,<sup>1</sup> Gary Wong,<sup>2</sup> Foday Daffae,<sup>4</sup> Abdul Kamara,<sup>4</sup> AiPing Wu,<sup>6</sup> Tai-Jiao Jiang,<sup>6</sup> Zhan Li,<sup>1</sup> Jie Huang,<sup>1</sup> Yu Sun,<sup>1</sup> Jun Qian,<sup>5</sup> Brima Kargbo,<sup>4</sup> Jia-Fu Jiang,<sup>1</sup> Hui Wang,<sup>1</sup> and Wu-Chun Cao<sup>1</sup>

<sup>1</sup>State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, China.

<sup>2</sup>Institute of Microbiology, Chinese Academy of Sciences, Beijing, China. <sup>3</sup>Beijing Institute of Biotechnology, Beijing, China.

<sup>4</sup>Sierra Leone Ministry of Health and Sanitation, Freetown, Sierra Leone. <sup>5</sup>Key Laboratory of Jilin Province for Zoonosis

Prevention and Control, Changchun, China. <sup>6</sup>Center for Systems Medicine, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; Suzhou Institute of Systems Medicine, Suzhou, China.

Sierra Leone was the most severely affected country in Western Africa during the 2013–2016 outbreak of Ebola virus disease (EVD). Previous genome surveillance studies have revealed the origin, diversity, and evolutionary dynamics of the Ebola virus (EBOV); however, the information regarding EBOV sequences is insufficient, especially the clinical outcomes, given that the correlation between the clinical outcomes and the genetic evolution of EBOV is still not clear. Here, we collected and curated a comprehensive data set that includes 514 EBOV genome sequences from patients with confirmed EVD (including 60 sequences not previously studied), >87.5% of which have residence information and definitive clinical outcomes. Phylogenetic reconstruction revealed 11 lineages of EBOV in Sierra Leone. The median-joining haplotype network showed that haplotypes that are associated with lethal outcomes tend to contribute more to the spread of the EBOV in Sierra Leone than those with live outcomes. Analyses of the spatial-temporal distribution unraveled the lineage-distinctive distribution patterns. Different viral lineages have different case fatality rates (CFRs) during the same stage of the outbreak, implying that several lineages featuring SNPs may correlate with increased/decreased CFRs. This study provides invaluable data sets of EBOV infection and highlights the potential SNPs for further in-depth investigation.

## Introduction

An unprecedented outbreak of Ebola virus disease (EVD) in Western Africa that began in 2013 was declared over in 2016; it resulted in 28,616 cases, with 11,310 deaths, as of March 29, 2016 (1). Sierra Leone, which was the most severely affected country in the area, reported its first confirmed EVD case on May 25, 2014. Deep sequencing of the viral genome from patient samples and phylogenetic studies have unraveled the evolution and transmission of the Ebola virus (EBOV) in Sierra Leone (2–4), Guinea (5–7), Liberia (8), and Mali (9). In our previous studies, we not only discovered a steady nucleotide substitution rate and diversified evolutionary lineages for EBOV, but also identified the transmission dynamic and intervention effectiveness of EVD in Sierra Leone (10), which may contribute to more closely understanding the epidemic characteristics of EVD and EBOV in Sierra Leone. However, the correlation between the clinical outcomes and the genetic evolution remain unclear.

In this study, we collected EBOV sequences in Sierra Leone, sequenced 60 EBOV genomes, and mapped the spatiotemporal distribution and clinical outcomes of EVD patients in Sierra Leone with the genomic evolution of EBOVs sampled from these patients, which showed both outcome-specific haplotype transmission features and lineage-specific case fatality rates (CFRs). These EBOV infection data and findings on key SNPs, which correlate with CFRs, may provide valuable information for further investigations of EBOV pathogenesis.

**Authorship note:** T. Li, H.W. Yao, D. Liu, H.G. Ren, Y. Hu, and D. Kargbo contributed equally to this work.

**Conflict of interest:** The authors have declared that no conflict of interest exists.

**Submitted:** May 4, 2016

**Accepted:** June 15, 2017

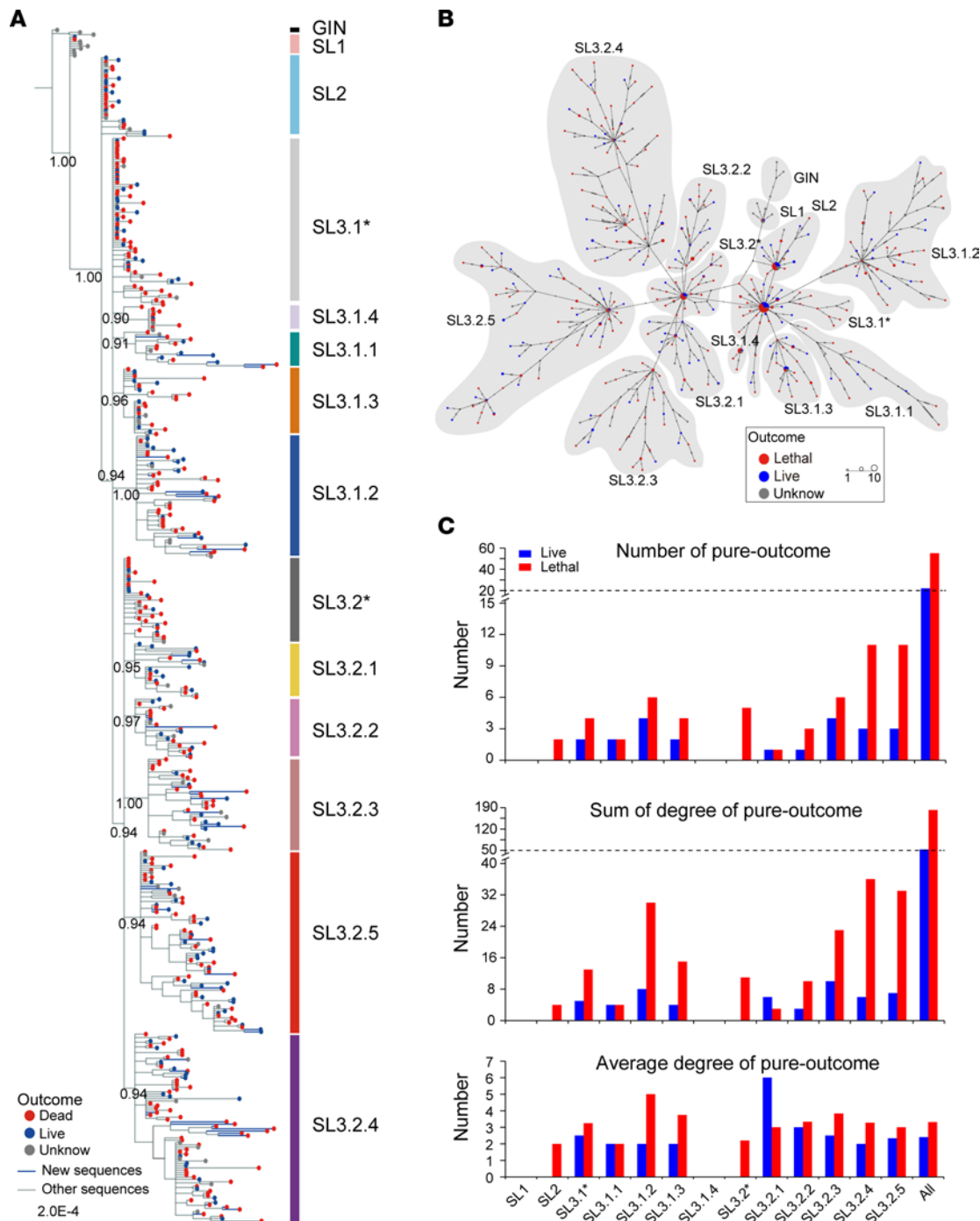
**Published:** July 20, 2017

**Reference information:**

JCI Insight. 2017;2(14):e88333.

<https://doi.org/10.1172/jci.insight.88333>

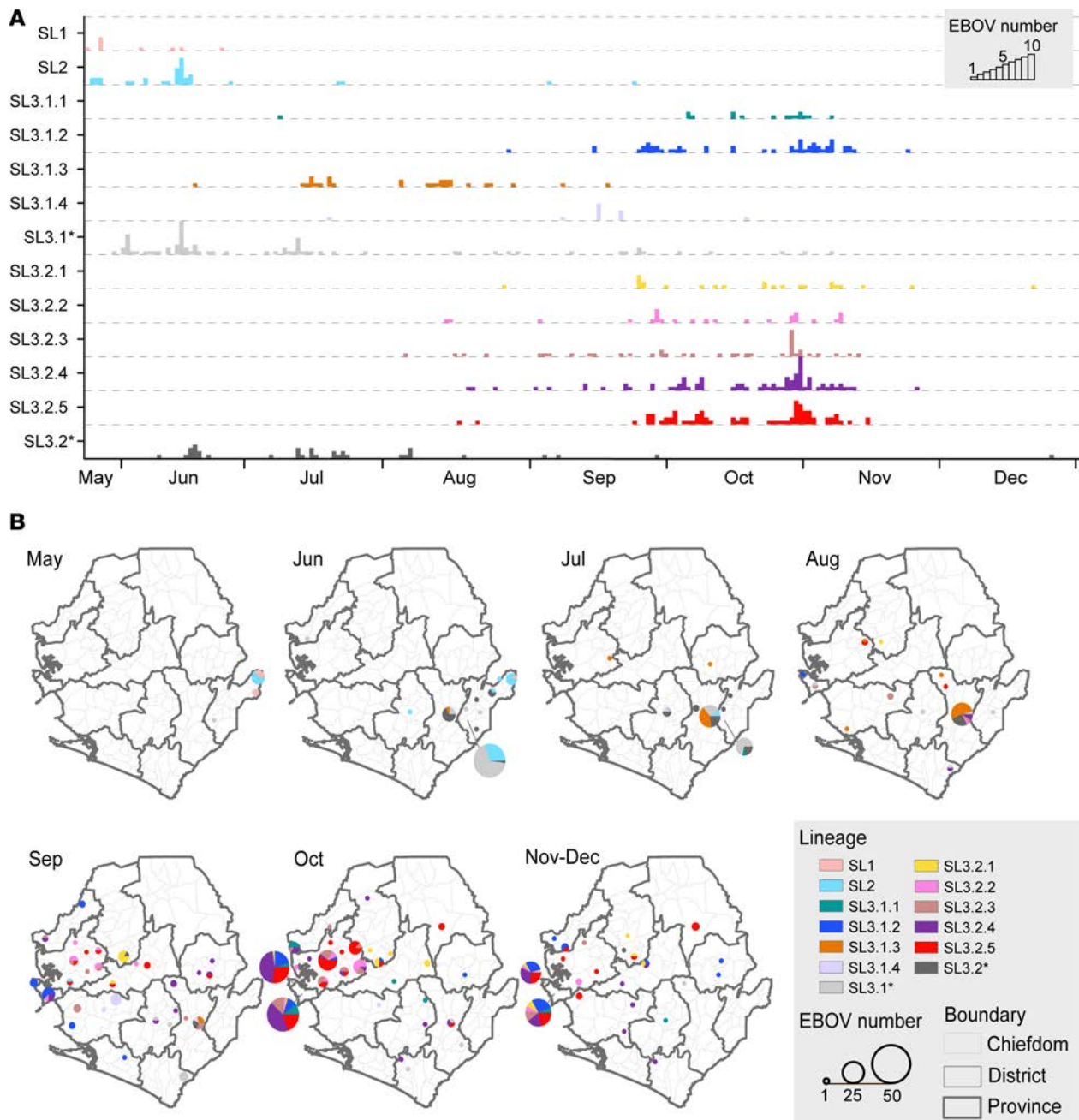
insight.88333.



**Figure 1. Phylogenetic analysis of EBOV, with clinical outcomes of patients in Sierra Leone.** (A) A Bayesian phylogenetic tree of 514 EBOV sequences from Sierra Leone and 3 early EBOV sequences from Guinea. The patient clinical outcome information is indicated by color. Posterior support for major nodes is shown. SL3.1\* means the other EBOV cases in SL3.1, except SL3.1.1, SL3.1.2, SL3.1.3, and SL3.1.4. SL3.2\* means the other EBOV cases in SL3.2, except SL3.2.1, SL3.2.2, SL3.2.3, SL3.2.4, and SL3.2.5. Posterior support for major nodes is shown. (B) Median-joining haplotype network for the EBOV sequences. The patient clinical outcomes are shown in different colors. (C) The number and degree of the pure-live and pure-lethal intermediate haplotypes in each lineage of EBOV.

## Results

In the current study, we performed a long-term investigation of EVD patients in Sierra Leone (9) and collected data on residence and definitive clinical outcome of the EVD cases. We also linked our previously sequenced and publicly released EBOV genomes to those patients (3–5, 7). In addition, we performed viral genome sequencing of 232 blood and swab samples and obtained 60 EBOV genomes (Supplemental Table 1; supplemental material available online with this article; <https://doi.org/10.1172/jci.insight.88333DS1>). We compiled a comprehensive data set that includes 514 patients in total, with 455 having follow-up definitive clinical outcome information, 501 having chiefdom-level residence information, and all 514 patients having precise virus sampling dates and genome sequences. The information covers the outbreak in Sierra Leonean chiefdoms over 8 months (May to December 2014), and the num-

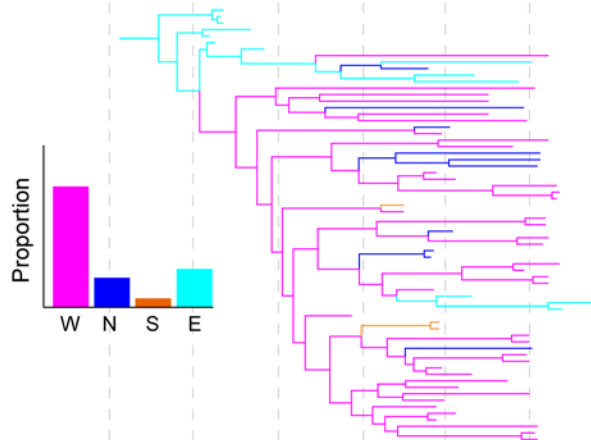


**Figure 2. Spatiotemporal distribution of each lineage of EBOV from Sierra Leone. (A)** Daily distribution of each lineage. The height of individual bars indicates the number of EBOV cases. **(B)** The spatiotemporal distribution of each lineage by chiefdom and month. Different colors show different lineages, and the size of the circles represent the number of EBOV sequences.

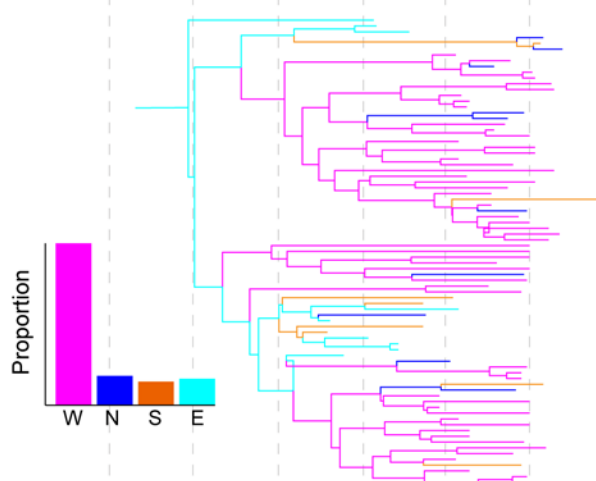
ber of daily patients are likely proportional to the number of confirmed EVD cases (Supplemental Figure 1). To our knowledge, this is the most comprehensive EBOV data set of patients to date from Sierra Leone. Since we do not have the detailed clinical outcome information for the sequences sampled from countries other than Sierra Leone, we focused our efforts on this data set in the following analysis.

Phylogenetic analysis of the 514 viral genome sequences shows that, over this period, 11 distinct lineages established themselves (Figure 1A), with 2 of these lineages (SL3.1.3 and SL3.1.4) being newly identified to our knowledge (3). In the median-joining haplotype network, 372 haplotypes have been discovered (Figure 1B). There were 205 of these haplotypes that resulted in 100% lethal outcomes (pure-lethal haplotypes) and 103 with 100% live outcomes (pure-survival haplotypes), among which 55 pure-lethal haplotypes and 22 pure-survival haplotypes act as intermediate nodes in the network. The average degrees of these pure-le-

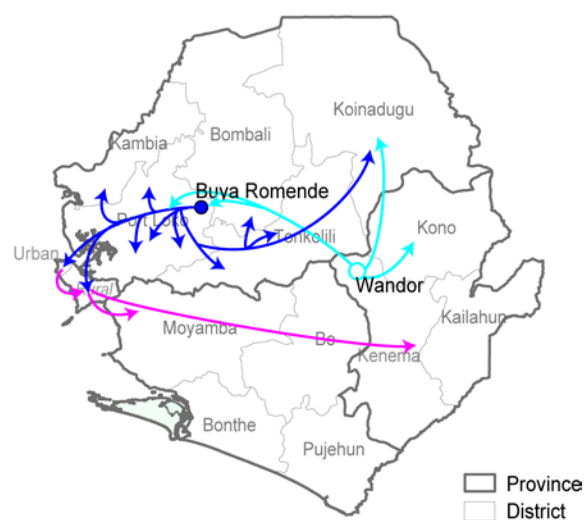
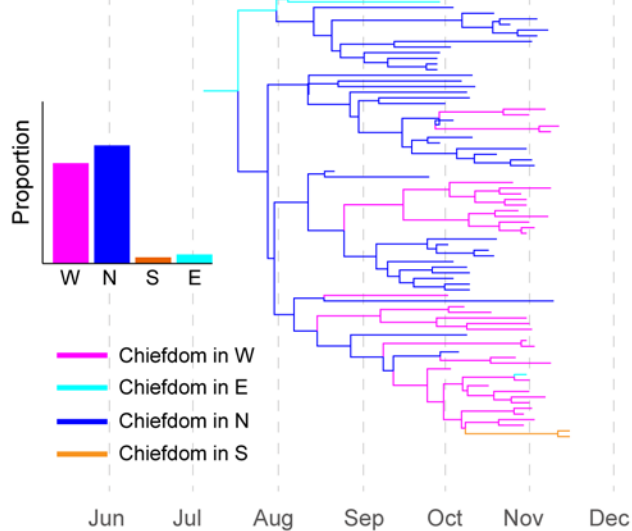
SL3.1.2



SL3.2.4



SL3.2.5





**Figure 3. Reconstructed phylogeographic linkages of lineages SL3.1.2, SL3.2.4, and SL3.2.5.** The phylogeographic linkage by chiefdom was constructed using BEAST. On the left side, histograms and branches with 4 colors in the MCC trees represent the proportion and chiefdoms in the Northern (N), Southern (S), Western (W), or Eastern (E) Provinces. On the right side, the transmission paths were constructed based on the trees and phylogeographic linkages with Bayes factor >3. The arrows in the 4 colors represent virus importation from N, S, W, or E Provinces of Sierra Leone, respectively.

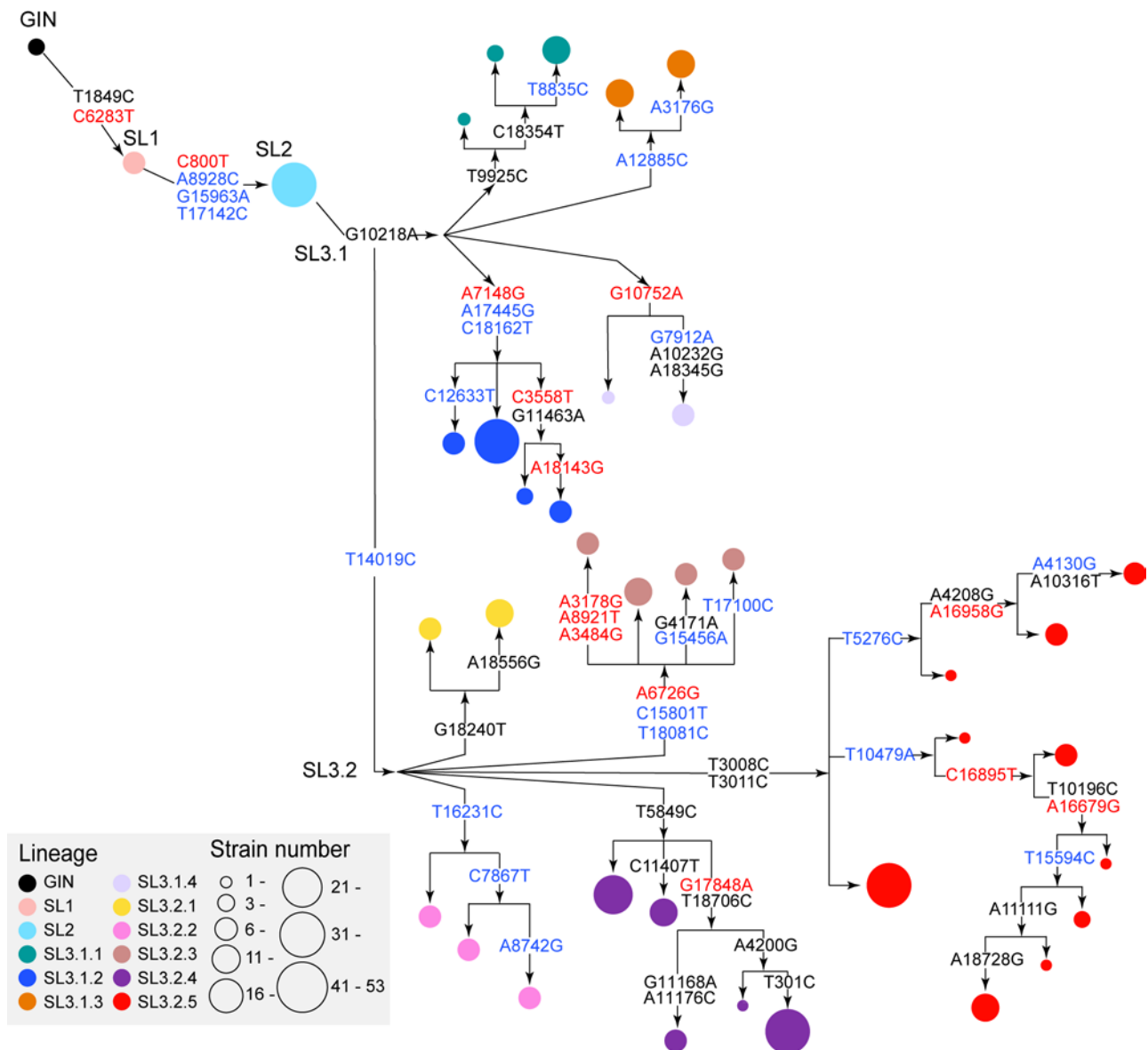
thal and pure-survival intermediate haplotypes are 3.31 and 2.41, respectively (Figure 1C). The higher average degree of the pure-lethal intermediate haplotypes implies that patients with lethal outcomes potentially infected more people than those with live outcomes, which may be related to the higher viremia in the nonsurvivors (11) or specific corpse handling and touching during funerals that contributed to the early spread of EVD in Sierra Leone (2).

From a spatial-temporal view, lineages SL1 and SL2 emerged in the early stages of the outbreak in the Eastern Province of Sierra Leone and were soon replaced by SL3 after June 2014 (Figure 2). Lineage SL3 was first detected in late May 2014 and became dominant in July 2014 through continuous expansion and diversification. From the diversified SL3 lineages, 3 lineages (SL3.1.2, SL3.2.4, and SL3.2.5) were shown to be more widespread (12, 23, and 18 chiefdoms, respectively) and to possibly infect more patients (53, 82, and 79 viruses, respectively, were identified from the patients). These 3 lineages would be the dominant viruses in the latter months of the outbreak in Sierra Leone in 2014 (Supplemental Figure 2). In October, the virus proportion was 10.4%, 29.5%, and 27.7% for SL3.1.2, SL3.2.4, and SL3.2.5, respectively, and in November the corresponding values were 25.0%, 21.4%, and 25.0% (Supplemental Figure 2). We also used the isolated provinces as traits to construct the median-joining haplotype network (Supplemental Figure 3), and we found that each lineage had distinctive transmission patterns. Generally, the EBOV was first introduced from Guinea into the Eastern Province of Sierra Leone, and it was sequentially transmitted to the Northern, Western, and Southern Provinces. This transmission pattern, revealed by phylogenetic analysis, was partially consistent with the expansion patterns of EVD cases in our previous study (10).

We then detailed the transmission paths of the 3 major EBOV lineages in Sierra Leone (Figure 3). Although all 3 lineages likely originated from the Eastern Province, they had distinctive transmission patterns. The spread of SL3.1.2 was with an advantage, in that its introduction to Freetown was at an earlier time than the other strains. In contrast, SL3.2.4 viruses were sequentially transmitted from the Eastern Province to adjacent chiefdoms, reaching Western Rural via Waterloo (a major transport hub). Then, SL3.2.4 spread to the Northern and Southern Provinces, in addition to frequent mutual Urban-Rural transmissions in the Western Province. Since then, SL3.2.4 became one of the major lineages in the Western Province, due to a great extent to the occupation of transport junctions. SL3.2.5 was the first among the 3 major lineages to reach the Northern Province, and the chiefdom of Buya Romende acted as the main transmission node. From there, SL3.2.5 spread to adjacent chiefdoms and established itself as the most prevalent lineage in the Northern Province.

To investigate SNPs that may contribute to the virus fitness among lineages, we reconstructed the evolutionary pathways and placed the SNPs on each node. In this study, the substitutions that clustered at least 8 isolates in the phylogenetic tree were defined as the possible fixed substitutions (Figure 4 and Supplemental Table 2). Among these possible fixed substitutions, the number of nonsynonymous mutations ( $n = 14$ ) was nearly 60% of that of synonymous ( $n = 23$ ) and noncoding ( $n = 23$ ) mutations. Notably, 5 of 14 nonsynonymous substitutions (at sites 16679, 16895, 16958, 17848, and 18143) were in the mRNA-capping methyltransferase domain of the L protein (12). Three amino acid mutations (sites 6283, 6726, and 7148) were in the GP1 domain of the GP protein, which affects the attachment to host cells (13). Specifically, A6726G in SL3.2.3 was also in the sequences of isolates in Guinea (6).

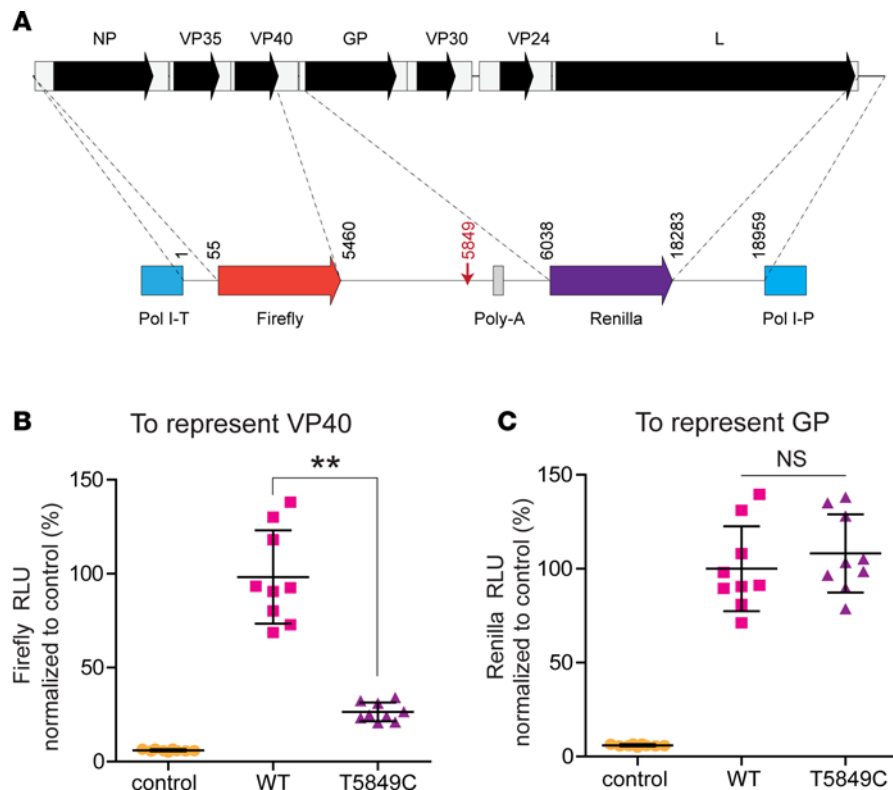
Then, we observed many featured substitutions in different lineages. Clearly, G10218A is a feature of lineage SL3.1, and T14019C is a feature of SL3.2 (Figure 4). As for lineage SL3.1.2, the dominant SNPs include 1 nonsynonymous mutation (A7148G, Ile371Val) in the GP gene/protein and 2 synonymous mutations (A17445G and C18162T) in the L gene. All viruses in SL3.2.4 shared SNP T5849C in the 3'-UTR of VP40, while all viruses in SL3.2.5 shared SNPs T3008C and T3011C in the 3'-UTR of NP. It is intriguing that the featured SNPs of both the dominant SL3.2.4 and SL3.2.5 are located in the 3'-UTR. Then, we noted that there was a high proportion of shared noncoding sites in the fixed mutant in the 2 lineages (7 of 8 in SL3.2.4 and 7 of 14 in SL3.2.5). These key substitutions possibly changed the expression level of relative proteins to improve the adaptation of the virus in Sierra Leone. Then, we employed an Ebola mini-genome reporting system to examine whether T-to-C substitutions at 5849 are functional (Figure 5A). As position 5849 is located



**Figure 4. Key SNPs and case fatality rates along lineage diversification.** All SNPs are called against a reference sequence taken in March 2014 (GenBank accession KJ660346). A substitution tree is shown to summarize evolutionary pathways. The substitutions that clustered at least 10 viruses in the phylogenetic tree are presented. Words in red, blue, and black indicate the nonsynonymous, synonymous, and noncoding SNPs, respectively.

between genes VP40 and GP, we applied the Firefly luciferase gene to denote the expression of VP40 and the Renilla luciferase gene to denote that of the GP gene. Our results showed that VP40 expression of the mutant with T-to-C substitutions at 5849 was significantly decreased ( $P < 0.01$ , Figure 5B). The expression of Renilla luciferase of the mutants with T-to-C substitutions did not change significantly (Figure 5C). These results clearly showed that the T-to-C substitutions at position 5849 affected the expression of the VP40 gene but not the GP gene. In addition, the results from a parallel study showed that linked T3008C and T3011C SNPs increase the transcription of NP through an EBOV mini-genome reporting system (14).

In general, the overall CFR for the 455 patients in our study was 66.2% (Figure 6), which is comparable to previous estimations (15, 16). The CFRs of lineages SL3.1 and SL3.2 were comparable to the overall CFR calculated in our study. The lineage with the highest CFR was determined to be SL3.1.4 (88.9%), and the lineage with the lowest CFR was SL3.2.1 (42.9%). Since neither of these lineages became dominantly circulated lineages, CFR is not the only indicator that predicts viral prevalence. Interestingly, the 2 major lineages circulating in Sierra Leone had different CFRs: SL3.2.4 had an increased CFR (73.0%) and SL3.2.5 showed a decreased CFR (58.3%). Featured SNPs in each lin-



**Figure 5. Examining T-to-C substitutions at position 5849 with Ebola mini-genome reporting system.** (A) Schematic of Ebola mini-genome reporting system. (B and C) Use of Firefly and Renilla luciferase to denote the expression of the VP40 gene and GP in the WT Ebola mini-genome system and mutants with T-to-C mutation at position 5849, respectively. The eGFP gene was used as the control. Pairwise comparisons show that the T-to-C mutation at 5849 significantly decreased the expression of VP40, but did not affect the GP expression (\*\* $P < 0.01$ ). Data represent mean  $\pm$  SD.

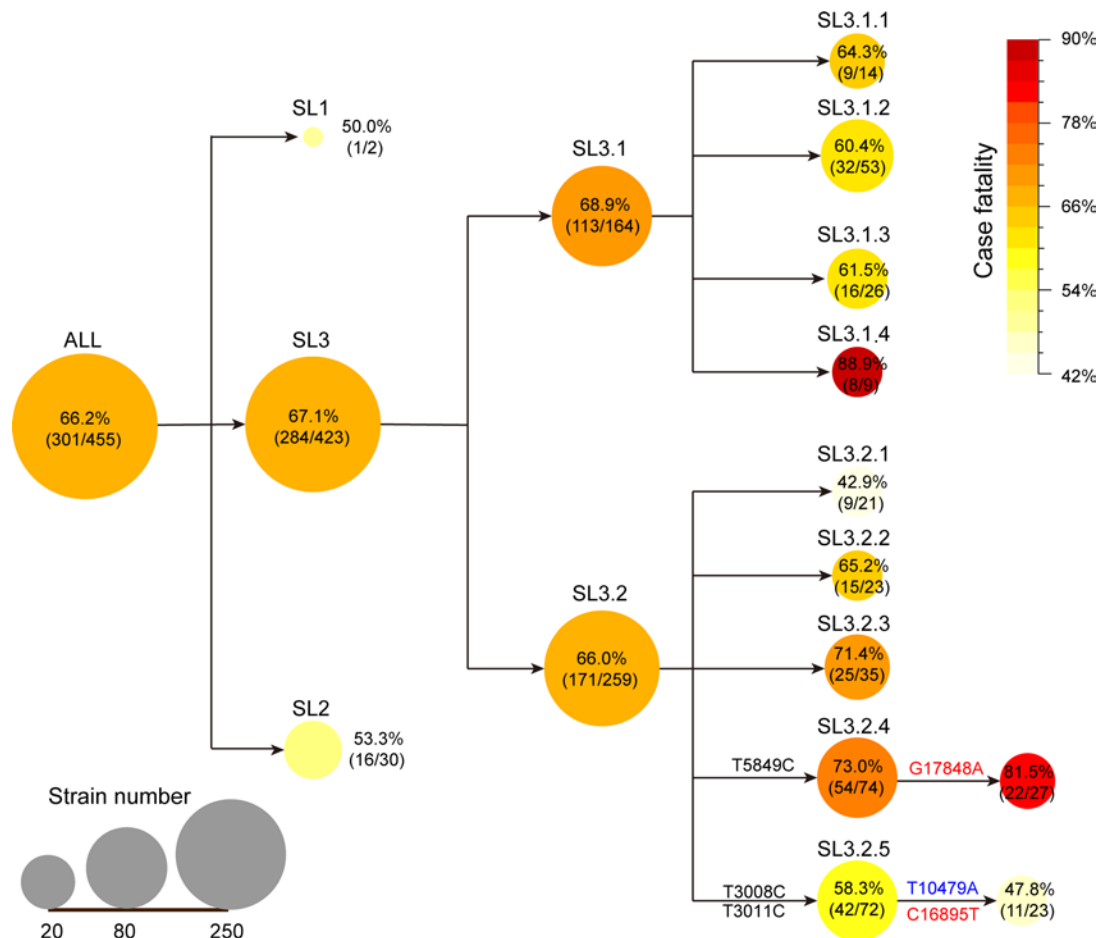
eage may have contributed to the differences among CFRs in those lineages. SL3.2.5 featuring SNPs 10479A and C16895T and SL3.2.1 featuring SNP G18240T were significantly correlated with a lower CFR ( $\chi^2$  test,  $P < 0.05$ , Figure 7A). It should be noted that from October to November 2014, when most of the SL3.2.4 viruses were identified (87.5%), the CFR of SL3.2.4 was significantly increased, which was very likely associated with SNPs T5849C and G17848A ( $\chi^2$  test,  $P < 0.05$ , Figure 7B). However, further verification is needed.

## Discussion

In this study, we collected and curated a comprehensive data set of 514 patients, including residence information, definitive clinical outcomes, and genomic sequences (of 60 not previously studied virus sequences). Then, we revealed that the lineages circulated in Sierra Leone from May to December 2014 exhibited distinct transmission patterns and CFRs. Phylogenetic reconstruction revealed 11 lineages of EBOV in Sierra Leone. The 3 major lineages that circulated within Sierra Leone from May to December 2014 exhibited distinct characteristics. Lineage SL3.1.2 arose from high population density, and its SNPs were very likely due to random genetic drift. SL3.2.4 represents a lineage with high levels of virus spread due to environmental and possibly viral factors. The increased CFR of SL3.2.4 indicates that the SNPs associated with this strain may affect virulence associated SNPs may affect virulence and viral transmissibility. SL3.2.5, which established itself amid SL3.1.2 and SL3.2.4 in Freetown, likely possesses virus adaptive/fitness features.

In addition, several lineages featuring SNPs, such as T3008C, T3011C, T10479A, C16895T, and T5849C, were mostly located at a noncoding region and correlated with the increased/decreased CFRs, implying that CFR-associated SNPs may affect virulence and/or viral transmissibility. All viruses in SL3.2.4 shared SNP T5849C in the 3'-UTR of VP40, a downregulated VP40 expression level, and a changed CFR, which is probably an adaptive feature and needs further investigation.





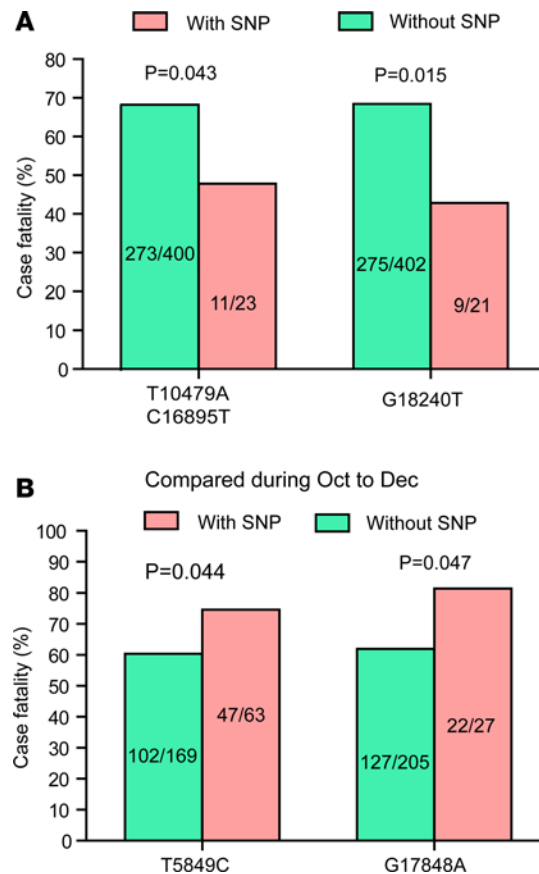
**Figure 6. Case fatality rates for different lineages in Sierra Leone.** The number of dead and total patients is shown within the circles. Colors from yellow to red indicate case fatality rates (CFRs). The size of the circle represents the total number of patients with outcome information in each lineage.

In conclusion, this study provides further knowledge about the correlation between clinical outcomes and the genetic evolution of EBOV in the 2014–2016 outbreak. These findings highlight potential SNPs of general interest for further in-depth clinical investigation and will further promote the basic study on EBOV pathogenesis.

## Methods

**Genome sequencing and assembly.** Next-generation sequencing was performed using the Illumina 4000. All of the sequenced reads were filtered using a custom-made program to remove the low quality and short reads. The genome sequences of the viruses were assembled by mapping the filtered reads to the 2014 EBOV consensus sequence using Bowtie2 v2.0.5., and the mutation site was manually checked with the original sequencing data. All of the EBOV genome sequences were aligned using Clustal Omega.

**Data collection and management.** All of the available EBOV genome nucleotide sequences and detail information regarding viruses in Sierra Leone were downloaded from the GenBank (<http://www.ncbi.nlm.nih.gov/GenBank/>) databases and derived from 4 appendix databases of literature (3–5, 7). If chiefdom information was not provided with viruses in the appendix databases (chiefdom is the third-level administrative unit in Sierra Leone, below province and district), we matched the “Patient ID or Lab ID” variable (for example, “G3764” of KM233063) in the appendix databases to the case database of the Sierra Leone Ministry of Health and Sanitation. After matching, we checked and complemented the chiefdom and outcome information for the viruses, which was recorded in the case database when it was missing in the appendix databases. Next, we excluded the viruses with a large portion of the information missing. Then, we formed a data set, which includes 3 EBOV sequences from Guinea (GenBank accession



**Figure 7. Significant SNPs associated with the CFRs.** (A) The test was performed in SL3 ( $n = 423$ ). (B) The test was performed in SL3 ( $n = 232$ ) from October to December. The number of dead and total patients in each column is shown ( $\chi^2$  test).

KJ660346, KJ660347, and KJ660348) (5), 454 published EBOV sequences from Sierra Leone, and 60 EBOV sequences from this study.

*Evolutionary and substitution inference.* All of the sequences from the data set were aligned using MAFFT v7.058 (17). Phylogenetic analyses were performed using MrBayes v3.2 (10 million generations), with the GTR model of nucleotide substitution and gamma-distributed rates among sites. The same sequence set was also analyzed in PopART v1.7.2 (<http://popart.otago.ac.nz>) using the median-joining haplotype network reconstruction method. We collected all of the SNPs of the other sequences and compared them to a reference sequence (KJ660346). SNPs, clustered on the MCC tree (Figure 1A) and shared by at least 10 EBOV sequences, were labeled on the branch of the substitution tree (Figure 3), which was manually illustrated from the MCC tree under the assumption that observed substitutions occurred only once. We used the Bayesian Markov chain Monte Carlo method as implemented in BEAST,

version 1.8, to make a time-scaled phylogenetic reconstruction. Bayes factor tests were performed to provide statistical support for potential transmission routes between different geographic locations using SPREAD v1.0.6 to elucidate the phylogeographic transmission of SL3.1.2, SL3.2.4, and SL3.2.5.

*Maps.* We obtained province, district, and chiefdom boundary data from the GADM database of Global Administrative Areas version 2.0 (<http://www.gadm.org>). All of the thematic maps shown in this study were created by ArcGIS 9.2 software (ESRI Inc.).

*Ebola mini-genome reporter assays.* The 293T cells seeded to approximately 80% confluency in 24-well plates were transfected with 0.18  $\mu$ g of each Ebola plasmid (L, NP, VP35, and VP30), 0.18  $\mu$ g of varied bicistronic reporter constructs WT and T5849C, and 0.4  $\mu$ g of pEGFP-N1 plasmid for normalization of each well with 4.0  $\mu$ l of polyethylenimine transfection reagent (Alfa Aesar). Transfected cells were washed once with PBS and then lysed with gentle rocking for 10 minutes at room temperature before being transferred to a 96-well white flat-bottomed polystyrene plate. Luminescence of the lysed cell was measured with the Dual-Luciferase Reporter Assay system kit (Promega) using a GLOMAX luminometer (Promega).

*Accession.* Next-generation sequencing data for all 60 genome sequences have been deposited in GenBank database (<http://www.ncbi.nlm.nih.gov/GenBank/>) under accession KU143775-KU143834.

*Statistics.* The  $\chi^2$  test was used to estimate the case fatality differences between groups with or without SNP. Student's  $t$  test was used to compare the expression levels in the Ebola mini-genome reporting system. All tests were 2 tailed, and  $P < 0.05$  was indicative of statistical significance. All statistical analyses were conducted by using SPSS 20.0.

*Study approval.* Blood samples from suspected individuals and oropharyngeal swab samples from corpses were collected for EVD testing and outbreak surveillance with a written informed consent waiver during the EVD outbreak, under an agreement between the Sierra Leone government and Chinese government. Activities were coordinated by the Emergency Operations Center in the charge of the Sierra Leone Ministry of Health and Sanitation and WHO. All the information regarding individual persons has been anonymized in the report.

## Author contributions

WCC, HW, JFJ, and BK designed the research study. YH, DK, YQD, HJL, LQF, AK, FD, YS, JQ, and JFJ collected samples and information. TL, HWY, DL, HGR, YH, YT, XL, KL, NZN, AW, TJJ, ZL, JH, JFJ, HW, and WCC performed the experiment and data analysis. TL, HWY, DL, HGR, GW, JFJ, HW, and WCC wrote the manuscript.

## Acknowledgments

This work was supported by the National Key R&D Program of China (grant 2016YFC1200301), National Key Basic Research Program 973 project (grant 2015CB554202), State Key Laboratory of Pathogen and BioSecurity Program (grant SKLPBS1408), National Key Technology R&D Program of the Ministry of Science and Technology (grant 1061400100275), and the Military Medical Innovation Project of the Academy of Military Medicine Science (2015CXJJ20).

Address correspondence to: Wu-Chun Cao, Hui Wang, or Jia-Fu Jiang, State Key Laboratory of Pathogen and Biosecurity, 20 Dong-Da Street, Fengtai District, Beijing 100071, China. Phone: 86.10.63896082; Email: caowc@bmi.ac.cn (W.C. Cao). Phone: 86.10.66948587; Email: geno0109@vip.sina.com (H. Wang). Phone: 86.10.63896082; Email: jiangjf2008@gmail.com (J.F. Jiang). Or to: Brima Kargbo, Sierra Leone Ministry of Health and Sanitation, Youyi Building, Bookfields, Freetown, Sierra Leone. Phone: 232.78515606; Email: brimakargbo@gmail.com.

1. Ebola Situation Reports. The World Health Organization. <http://www.who.int/csr/don/archive/disease/ebola/en/>. Accessed June 15, 2017.
2. Gire SK, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science*. 2014;345(6202):1369–1372.
3. Tong YG, et al. Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Leone. *Nature*. 2015;524(7563):93–96.
4. Park DJ, et al. Ebola virus epidemiology, transmission, and evolution during seven months in Sierra Leone. *Cell*. 2015;161(7):1516–1526.
5. Baize S, et al. Emergence of Zaire Ebola virus disease in Guinea. *N Engl J Med*. 2014;371(15):1418–1425.
6. Simon-Loriere E, et al. Distinct lineages of Ebola virus in Guinea during the 2014 West African epidemic. *Nature*. 2015;524(7563):102–104.
7. Carroll MW, et al. Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. *Nature*. 2015;524(7563):97–101.
8. Ladner JT, et al. Evolution and spread of Ebola virus in Liberia, 2014–2015. *Cell Host Microbe*. 2015;18(6):659–669.
9. Hoenen T, et al. Virology. Mutation rate and genotype variation of Ebola virus from Mali case sequences. *Science*. 2015;348(6230):117–119.
10. Fang LQ, et al. Transmission dynamics of Ebola virus disease and intervention effectiveness in Sierra Leone. *Proc Natl Acad Sci USA*. 2016;113(16):4488–4493.
11. de La Vega MA, et al. Ebola viral load at diagnosis associates with patient outcome and outbreak evolution. *J Clin Invest*. 2015;125(12):4421–4428.
12. Cong Q, Pei J, Grishin NV. Predictive and comparative analysis of Ebolavirus proteins. *Cell Cycle*. 2015;14(17):2785–2797.
13. Lee JE, Fusco ML, Hessel AJ, Oswald WB, Burton DR, Saphire EO. Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. *Nature*. 2008;454(7201):177–182.
14. Ni M, et al. Intra-host dynamics of Ebola virus during 2014. *Nat Microbiol*. 2016;1(11):16151.
15. Schieffelin JS, et al. Clinical illness and outcomes in patients with Ebola in Sierra Leone. *N Engl J Med*. 2014;371(22):2092–2100.
16. Lu HJ, et al. Ebola virus outbreak investigation, Sierra Leone, September 28–November 11, 2014. *Emerging Infect Dis*. 2015;21(11):1921–1927.
17. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–780.