# LETTER

# Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Leone

Yi-Gang Tong[1]*, Wei-Feng Shi[2]*, Di Liu[3]*, Jun Qian[5]*, Long Liang[1]*, Xiao-Chen Bo[4]*, Jun Liu[6]*, Hong-Guang Ren[1]*, Hang Fan[1]*, Ming Ni[4]*, Yang Sun[5]*, Yuan Jin[1], Yue Teng[1], Zhen Li[4], David Kargbo[7], Foday Dafae[7], Alex Kanu[8], Cheng-Chao Chen[9], Zhi-Heng Lan[9], Hui Jiang[9], Yang Luo[10], Hui-Jun Lu[5], Xiao-Guang Zhang[6], Fan Yang[11], Yi Hu[1], Yu-Xi Cao[6], Yong-Qiang Deng[1], Hao-Xiang Su[11], Yu Sun[1], Wen-Sen Liu[5], Zhuang Wang[1], Cheng-Yu Wang[5], Zhao-Yang Bu[5], Zhen-Dong Guo[5], Liu-Bo Zhang[12], Wei-Min Nie[13], Chang-Qing Bai[14], Chun-Hua Sun[1], Xiao-Ping An[1], Pei-Song Xu[4], Xiang-Li-Lan Zhang[1], Yong Huang[1], Zhi-Qiang Mi[1], Dong Yu[1], Hong-Wu Yao[1], Yong Feng[15], Zhi-Ping Xia[5], Xue-Xing Zheng[5], Song-Tao Yang[5], Bing Lu[1], Jia-Fu Jiang[1], Brima Kargbo[7], Fu-Chu He[16], George F. Gao[3,6,17], Wu-Chun Cao[1] & The China Mobile Laboratory Testing Team in Sierra Leone†

A novel Ebola virus (EBOV) first identified in March 2014 has infected more than 25,000 people in West Africa, resulting in more than 10,000 deaths[1,2]. Preliminary analyses of genome sequences of 81 EBOV collected from March to June 2014 from Guinea and Sierra Leone suggest that the 2014 EBOV originated from an independent transmission event from its natural reservoir[3] followed by sustained human-to-human infections[4]. It has been reported that the EBOV genome variation might have an effect on the efficacy of sequence-based virus detection and candidate therapeutics[5,6]. However, only limited viral information has been available since July 2014, when the outbreak entered a rapid growth phase[7]. Here we describe 175 full-length EBOV genome sequences from five severely stricken districts in Sierra Leone from 28 September to 11 November 2014. We found that the 2014 EBOV has become more phylogenetically and genetically diverse from July to November 2014, characterized by the emergence of multiple novel lineages. The substitution rate for the 2014 EBOV was estimated to be $1.23 \times 10^{-3}$ substitutions per site per year (95% highest posterior density interval, $1.04 \times 10^{-3}$ to $1.41 \times 10^{-3}$ substitutions per site per year), approximating to that observed between previous EBOV outbreaks. The sharp increase in genetic diversity of the 2014 EBOV warrants extensive EBOV surveillance in Sierra Leone, Guinea and Liberia to better understand the viral evolution and transmission dynamics of the ongoing outbreak. These data will facilitate the international efforts to develop vaccines and therapeutics.

A large-scale Ebola viral disease (EVD) outbreak has been ongoing in Western Africa for nearly a year, with more than 23,000 reported cases[1]. Previous findings have shown that the causative agent is a novel Ebola virus (EBOV)[2]. Among the three West African countries with widespread and intense EBOV transmission, Sierra Leone reported the largest number of confirmed cases, approximately 58% of the total confirmed EBOV infection cases. To help Sierra Leone fight against EVD, the Chinese government dispatched the China Mobile Laboratory Testing Team (CMLTT) in September upon request of the Sierra Leone government. The CMLTT, equipped with medical experts who specialize in laboratory testing, epidemiology, and running a holding and treatment centre, has kept working at the Sierra Leone-China Friendship Hospital at Jui Town (represented as a red

star in Fig. 1a) of Western Area, approximately 30 km southeast of Freetown, the capital city of Sierra Leone. All the activities of the CMLTT were coordinated by the Emergency Operations Center jointly established by the Ministry of Health and Sanitation of Sierra Leone and the World Health Organization (WHO).

To fight against this novel EBOV, Gire and colleagues systematically analysed 81 EBOV genomes from Guinea ($n = 3$)[2] and Sierra Leone ($n = 78$)[4] collected from the early stage of the 2014 EBOV outbreak, revealing the origin, transmission, and rapid accumulation of genetic variation of the 2014 EBOV. However, only a few additional full-length EBOV genome sequences were published since July 2014, when the outbreak entered a rapid growth phase driven by sustained human-to-human transmission[4]. From 28 September to 11 November 2014, a total of 823 samples were tested to be EBOV-positive using reverse transcription-PCR (RT–PCR) by the CMLTT, among which 175 full-length genomes were successfully sequenced with each from an individual EVD patient (Fig. 1a and Supplementary Table 1). These 175 samples were obtained from five severely stricken districts in Sierra Leone, including 47 from Western Urban, 67 from Western Rural, 47 from Port Loko, 5 from Kambia, and 9 from Bombali (Fig. 1a). In detail, approximately one fifth of the EBOV-positive samples for each region were sequenced, 19.5% for Western Urban, 21.2% for Western Rural, 22.1% for Port Loko, and 16.1% for Kambia. Regarding Bombali, 9 out of 17 (52.9%) strains were sequenced. Therefore, our sequenced genomes were roughly proportional to the prevalence in different regions.

Phylogenetic analysis of all available full-length EBOV genome sequences from Sierra Leone ($n = 253$) and Guinea ($n = 3$) from 2014 was performed using MrBayes[8] in which the three Guinean strains were designated as root[4,9]. Our phylogenetic analysis showed that the 2014 EBOV increased in diversity at least through October after its initial introduction into Sierra Leone (Fig. 1b and Extended Data Fig. 1). Apart from the previously described lineages SL1 and SL2[4], the SL3 lineage has evolved into two major lineages, SL3.1 and SL3.2 in June in eastern Sierra Leone, both of which were then transmitted to western Sierra Leone. The majority of the EBOV collected from late September to mid-November fell into lineage SL3.2, with a few belonging to lineage SL3.1. However, none of them belonged to lineages SL1 and SL2. In particular, the EBOV sequenced by us could

[1]State Key Laboratory of Pathogen and Biosecurity, Beijing 100071, China. [2]Institute of Pathogen Biology, Taishan Medical College, Taian 271000, China. [3]Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China. [4]Beijing Key Laboratory of New Molecular Diagnostics Technology, Beijing 100850, China. [5]Key Laboratory of Jilin Province for Zoonosis Prevention and Control, Changchun 130122, China. [6]Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China. [7]Sierra Leone Ministry of Health and Sanitation, Freetown, Sierra Leone. [8]Sierra Leone-China Friendship Hospital, Freetown, Sierra Leone. [9]BGI-Shenzhen, Shenzhen 518083, China. [10]Wellcome Trust Sanger Institute, Cambridge CB10 1SA, UK. [11]Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China. [12]Institute of Environmental Health and Related Product Safety, Chinese Center for Disease Control and Prevention, Beijing 100021, China. [13]The No. 302 Hospital, Beijing 100039, China. [14]The No. 307 Hospital, Beijing 100071, China. [15]Department of international cooperation, National Health and Family Planning Commission, Beijing 100044, China. [16]State Key Laboratory of Proteomics, Beijing 102206, China. [17]Chinese Center for Disease Control and Prevention, Beijing 102206, China.
*These authors contributed equally to this work.
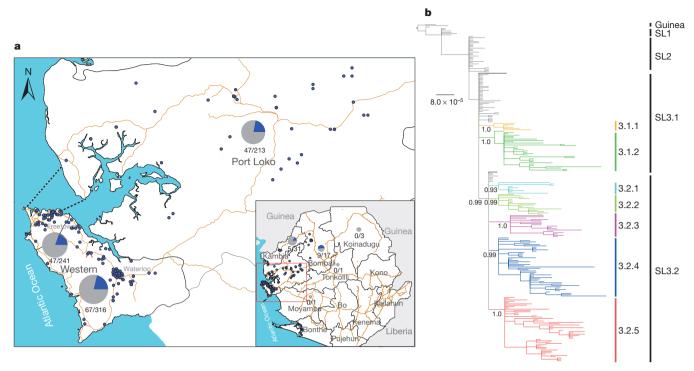†Lists of participants and their affiliations appear at the end of the paper

**Figure 1 | Geographical distribution and phylogenetic analysis of the 2014 EBOV from Sierra Leone. a**, Geographical distribution of the 823 EBOV positive samples and the 175 newly sequenced genomes (represented as blue dots). In the panel, main roads and waterways are showed as yellow lines and black dash lines, respectively. **b**, A Bayesian phylogenetic tree of the 2014 EBOV. The 175 newly sequenced viruses in this study are shown in colours, and others are shown in grey. The seven novel lineages designated in the present are highlighted. Posterior support for major nodes is shown.

be classified into seven novel independent sublineages based on the phylogenetic topology, two sublineages belonging to SL3.1 (SL3.1.1 and SL3.1.2) and five belonging to SL3.2 (SL3.2.1 to SL3.2.5) (Fig. 1b). Phylogenetic tree constructed using the maximum likelihood method showed a similar topology (Extended Data Fig. 2). Therefore, the 2014 EBOV has become highly diverse in its first year along with its spread in Sierra Leone.

To explore the spatiotemporal relationships of the EBOV in western Sierra Leone, we performed a phylogeographic analysis using BEAST[10] (Fig. 2 and Extended Data Fig. 3). In this analysis, only 22 out of the 78 sequences previously published by Gire *et al.* (ref. 4) were included in our analysis to reduce the computation load. To this end, we selected representative sequences from the previously described lineages GIN, SL1, SL2 and SL3, ensuring that there is at least one sequence for every sampling date. From a time point of view, all of the novel sublineages probably emerged before August (Fig. 2). In addition, multiple lineages were co-circulating in a single town/district. All of the seven sublineages were identified in Waterloo, indicating the highest phylogenetic diversity in this region. Viruses from Freetown belonged to six of the seven sublineages, with sublineage 3.2.3 undetected. Five novel sublineages have also been found in Maforki Chiefdom of Port Loko.

The spatiotemporal linkage of our sequenced EBOV genomes is further shown in Fig. 3a. First, viruses from Freetown and Waterloo, the capital and the traffic hub, are estimated to be spatiotemporally related, as also observed in sublineages 3.1.1, 3.1.2 and 3.2.4 (Fig. 2), indicating that frequent transmission events might have occurred between the two regions. Second, this network reveals that viral transmission events have also occurred between the three major sites (Freetown, Waterloo and Maforki Chiefdom) and their surrounding regions. Third, our results also suggest spatiotemporal connections of EBOV between Waterloo and Port Loko, Kambia, and Bombali, respectively, as exemplified in sublineages 3.2.1 and 3.2.5 (Fig. 2). Based on the higher transmission rates of Waterloo, Freetown and Maforki Chiefdom, intensive EBOV surveillance in
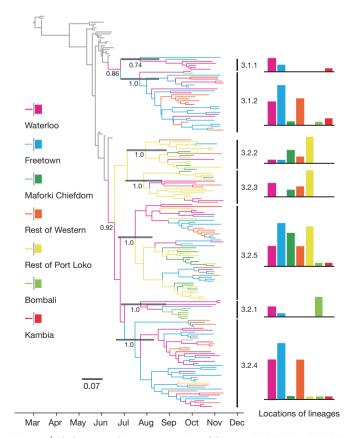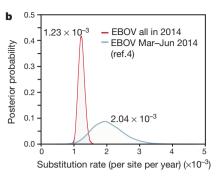


**Figure 2 | Phylogeographic reconstruction of the 2014 EBOV using BEAST.** In the left panel, the novel 175 EBOV genome sequences were coloured by geographic regions. The transition of different colours represents a potential transmission event. In the right panel, the number of sequences from different geographic regions in each lineage is summarized.
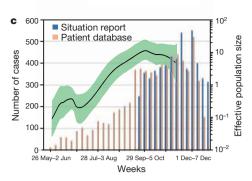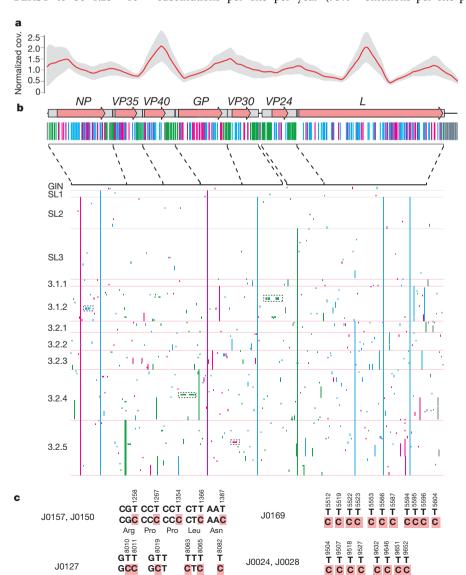
**Figure 3 | Reconstructed phylogeographic linkage, substitution rate, and effective population size of the 2014 EBOV in western Sierra Leone from September to November 2014. a**, The phylogeographic linkage constructed using BEAST. Thickness of lines represents the relative transmission rate between two regions. The size of each node is proportional to the sum of the relative rates of the region with Bayes factor >3. **b**, Substitution rates of the 2014 EBOV. The red line represents the substitution rate estimated using all the 2014 EBOV samples. Estimations of Gire and colleagues were repeated by us and shown as the blue line. **c**, Gaussian Markov random field Bayesian skyride reconstruction of the 2014 EBOV. Bar chart shows the numbers of confirmed cases of EBOV infection and patients. Smooth black line shows the effective population size. Adapted, with permission, from Ebola response roadmap - Situation report, Figure 3. http://www.who.int/csr/disease/ebola/situation-reports/en (accessed 1 April 2015)[1].

the three regions should be helpful for the prevention and control of the EVD outbreak in Western Sierra Leone.

The substitution rate for all of the 2014 EBOV was estimated using BEAST to be $1.23 \times 10^{-3}$ substitutions per site per year (95% highest posterior density interval, $1.04 \times 10^{-3}$ to $1.41 \times 10^{-3}$ substitutions per site per year) (Fig. 3b). Our estimate was similar to those between previous EBOV outbreaks, approximately $1.00 \times 10^{-3}$ substitutions per site per year[4,11–14]. This suggests that, over a longer



**Figure 4 | Genomic variations of the 2014 EBOV. a**, Sequence depth across sequenced genomes. The *x* axis represents the virus genome structure, and the *y* axis represents the normalized average depth. One unit equals approximately 1,400 coverage per site. The mean depth is shown using the red line and the standard deviation is shown in shade. **b**, Substitutions of the 2014 EBOV. Only positions with substitutions are shown. Different lineages are separated by lines. Different types of substitutions are indicated using different colours: cyan for synonymous (S), magenta for non-synonymous (NS), green for un-translated regions (UTR), and grey for intergenic regions (IG). **c**, All the serial T > C substitutions are found within a range less than 150 bp. Substitutions within coding regions are shown in codons.

time interval, EBOV is still undergoing evolution at a relatively constant rate.

The estimated population size of the 2014 EBOV from Sierra Leone steadily increased from July to early October, and then entered a plateau period (Fig. 3c). This therefore implies that the effective population size of the 2014 EBOV became stable in October, which was also broadly consistent with the weekly change of numbers of confirmed EBOV infection cases and EVD patients in Sierra Leone (Fig. 3c)[1]. The doubling time estimated using BEAST was 22.1 days (95% confidence interval, 18.9–25.59 days), which was comparable to that calculated using the epidemiological data from Sierra Leone, with the mean value of 18.9 days.

We then investigated the molecular characterization of the novel EBOV genome. Raw reads of each genome were mapped to the reference genome (KJ660346.2). The average normalized coverage was approximately 1,400-fold (Fig. 4a). 341 single nucleotide polymorphisms (SNPs) have been previously identified between the 2014 outbreak EBOV and previous EBOV[4], and 440 SNPs were identified in our sequenced genomes. The substitutions in the 175 newly sequenced EBOV genomes were summarized among different lineages (Fig. 4b and Supplementary Table 2). Approximately a quarter of the identified substitutions were non-synonymous, and half of them were synonymous (Extended Data Fig. 5). Some of the SNPs were lineage-specific and could be used as markers to distinguish different lineages (Fig. 4b and Supplementary Table 2). For example, substitutions A7148G and A17445G were only found in sublineage 3.1.2, whereas sublineage 3.2.4 possessed a specific T5849C substitution. The $T > C$ substitutions that occurred in the 3′ UTR region of $NP$ gene (at genome positions 3008 and 3011) were specific to sublineage 3.2.5. In particular, the $T > C$ substitution at position 14019 occurred in all sequences of lineage 3.2, which was first described in this study. Moreover, seven previously reported substitutions (at positions 800, 1849, 6283, 8928, 10218, 15963, 17142)[4] were always present in the novel lineages from June to November 2014 and became the dominant allele in the population, suggesting that they have been fixed. These substitutions included two non-synonymous substitutions (C800T in the $NP$ gene and C6283T in the $GP$ gene), four synonymous substitutions, and one in the non-coding regions.

Interestingly, we observed several serial $T > C$ substitutions in six newly sequenced EBOV genomes, which occurred within a genome region of 150 base pairs in length (Fig. 4c and Extended Data Fig. 4). The serial $T > C$ substitutions were further confirmed by Sanger sequencing after PCR amplification (Extended Data Table 1). Such serial substitutions were found in four different regions of six strains belonging to three different lineages, two of which were in coding regions and the other two were in non-coding regions. However, the emergence mechanism of such serial $T > C$ substitutions and their potential biological functions warrant further investigation.

In summary, our findings highlighted the increasing genetic diversity and transmission dynamics of the 2014 EBOV, with an evolutionary rate estimated to be similar to that between previous EBOV outbreaks. This information provided an insight into the viral evolution and transmission dynamics, which would facilitate the prevention and control of EBOV in Sierra Leone and would also guide research on vaccines and therapeutic targets.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. World Health Organization. Ebola response roadmap - Situation report. http://www.who.int/csr/disease/ebola/situation-reports/en (accessed 1 April 2015).
2. Baize, S et al. Emergence of Zaire Ebola virus disease in Guinea. N. Engl. J. Med. **371,** 1418–1425 (2014).
3. Leroy, E. M. et al. Fruit bats as reservoirs of Ebola virus. Nature **438,** 575–576 (2005).
4. Gire, S. K. et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science **345,** 1369–1372 (2014).
5. Kugelman, J. R. et al. Evaluation of the potential impact of Ebola virus genomic drift on the efficacy of sequence-based candidate therapeutics. MBio **6,** e02227–14 (2015).
6. Feldmann, H. et al. Ebola virus: from discovery to vaccine. Nature Rev. Immunol **3,** 677–685 (2003).
7. WHO Ebola Response Team. Ebola virus disease in West Africa–the first 9 months of the epidemic and forward projections. N. Engl. J. Med. **371,** 1481–1495 (2014).
8. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17,** 754–755 (2001).
9. Dudas G., Rambaut A. Phylogenetic analysis of Guinea 2014 EBOV Ebolavirus outbreak. PLoS Curr. http://dx.doi.org/10.1371/currents.outbreaks.84eefe5ce43ec9dc0bf0670f7b8b417d (2014).
10. Drummond, A. J. &. Rambaut. A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. **7,** 214 (2007).
11. Jenkins, G. M. et al. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J. Mol. Evol. **54,** 156–165 (2002).
12. Calvignac-Spencer S., et al. Clock rooting further demonstrates that Guinea 2014 EBOV is a member of the Zaire lineage. PLoS Curr. http://dx.doi.org/10.1371/currents.outbreaks.c0e035c86d721668a6ad7353f7f6fe86 (2014).
13. Carroll, S. A. et al. Molecular evolution of viruses of the family Filoviridae based on 97 whole-genome sequences. J. Virol. **87,** 2608–2616 (2013).
14. Li, Y. H. &. Chen. S. P. Evolutionary history of Ebola virus. Epidemiol. Infect. **142,** 1138–1145 (2014).

**Author Contributions** The manuscript was written by Y.-G.T., W.-F.S., D.L., G.F.G. and W.-C.C. Samples were collected by J.Q., D.K., F.D., A.K., B.K., Y.S., H.-J.L., X.-G.Z., F.Y., Y.H., Y.-X.C., Y.-Q.D., H.-X.S., Y.S., W.-S.L., Z.W., C.-Y.W., Z.-Y.B., Z.-D.G., L.-B.Z., W.-M.N., C.-Q.B., C.-H.S., Y.F., Z.-P.X., X.-X.Z., S.-T.Y. and B.L. Experiment and data analysis were performed by Y.-G.T., W.-F.S., D.L., H.F., M.N., H.-G.R., J.L., Y.J., Y.T., Z.L., C.-C.C., Z.-H.L., H.J., Y.L., X.-P.A., P.-S.X., X.-L.-LZ., Y.H., Z.-Q.M., D.Y., H.-W.Y., J.-F.J., X.-C.B., L.L., F.-C.H. and W.-C.C. The study was designed by B.K., X.-C.B., L.L., J.Q., F.-C.H., G.F.G. and W.-C.C.

**The China Mobile Laboratory Testing Team in Sierra Leone**

Yi-Gang Tong[1], Jun Qian[2], Yang Sun[2], Hui-Jun Lu[2], Xiao-Guang Zhang[3], Fan Yang[4], Yi Hu[1], Yu-Xi Cao[3], Yong-Qiang Deng[1], Hao-Xiang Su[4], Yu Sun[1], Wen-Sen Liu[2], Zhuang Wang[1], Cheng-Yu Wang[2], Zhao-Yang Bu[2], Zhen-Dong Guo[2], Liu-Bo Zhang[5], Wei-Min Nie[6], Chang-Qing Bai[7], Chun-Hua Sun[1], Yong Feng[8], Jia-Fu Jiang[1] & George F. Gao[3,9,10]

[1]State Key Laboratory of Pathogen and Biosecurity, Beijing 100071, China. [2]Key Laboratory of Jilin Province for Zoonosis Prevention and Control, Changchun 130122, China. [3]Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China. [4]Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China. [5]Institute of Environmental Health and Related Product Safety, Chinese Center for Disease Control and Prevention, Beijing 100021, China. [6]The No. 302 Hospital, Beijing 100039, China. [7]The No. 307 Hospital, Beijing 100071, China. [8]Department of international cooperation, National Health and Family Planning Commission, Beijing 100044, China. [9]Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China. [10]Chinese Center for Disease Control and Prevention, Beijing 102206, China.

## METHODS

**Ethics statement.** This work was conducted as part of the surveillance and public health response to contain the EVD outbreak in Sierra Leone. Blood samples from suspected individuals and oropharyngeal swab samples from corpses were collected for EVD testing and outbreak surveillance with a waiver to provide a written informed consent during the EVD outbreak under the agreement between the Sierra Leone government and Chinese government. The activities were coordinated by the Emergency Operations Centre in the charge of Sierra Leone Ministry of Health and Sanitation and WHO. All the information regarding individual persons has been anonymized in the report.

**Genome sequencing and assembly.** RNA samples extracted from whole blood from 175 EVD patients were reverse transcribed to cDNA. PCR amplifications were performed with EBOV-specific primer pairs with overlaps. Amplicons from one patient were pooled for library preparation. Next generation sequencing (NGS) was performed using the BGISEQ-100 (Ion Proton) platform. All the sequenced reads were filtered 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 Roche 454 Newbler version 2.9 (Roche), and the mutation site was manually checked with original sequencing data.

**Phylogenetic and phylogeographic reconstruction.** All previously published EBOV genome sequences and our newly released 175 sequences were aligned using MAFFT v7.058[15]. Phylogenetic analyses were performed using MrBayes[8] v3.2 (10 million generations) and RAxML v8.1.6 (1000 bootstrap replicates), with the GTR model of nucleotide substitution and $\gamma$-distributed rates among sites. Phylogeographic reconstruction of the 2014 EBOV was estimated using BEAST v1.8.0[10], with a continuous time Markov Chain (CTMC) over discrete sampling locations. The 175 newly sequenced samples in this paper were grouped into 7 regions (Waterloo, Freetown, Rest of Western, Maforki Chiefdom, Rest of Port Loko, Bombali and Kambia). Bayesian Markov chain Monte Carlo analysis was run for 100 million steps, 10% of which 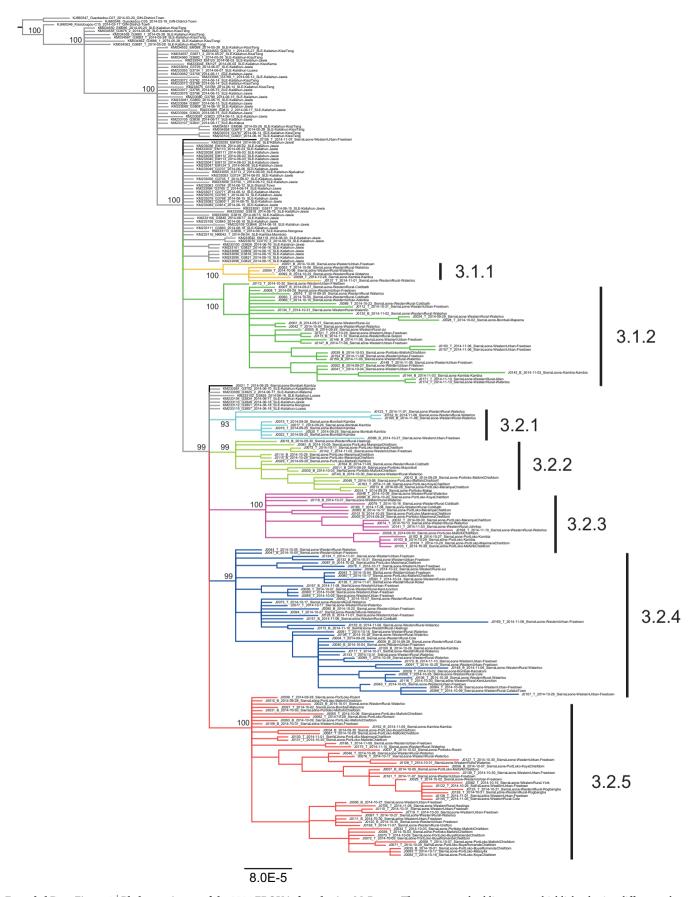were removed as burn-in and sampled every 10,000 steps. Bayes factor tests were performed to provide statistical support for potential transmission routes between different geographic locations using SPREAD v1.0.6[16]. Bayes factors for rates were derived from a Bayesian stochastic search variable selection procedure. The phylogeographic linkage was constructed by routes with Bayes factor values >3.

**Substitution rates and population dynamics.** The substitution rates were estimated using Bayesian Markov chain Monte Carlo (MCMC) as implemented in BEAST v1.8.0. In this analysis, two data sets were compiled, with one including all the 2014 EBOV sequences and the other including sequences from September to November, 2014. We performed two independent runs for 100 million generations, sampling every 10,000 steps. In addition, to accurately estimate the substitution rate, we repeated this analysis using a previously described data set using the same parameters. Population dynamics of the 2014 EBOV in Sierra Leone was estimated using a flexible non-parametric Bayesian skyride model[17] incorporated in BEAST v1.8.0, with the HKY+$\Gamma$ model and a strict molecular clock.
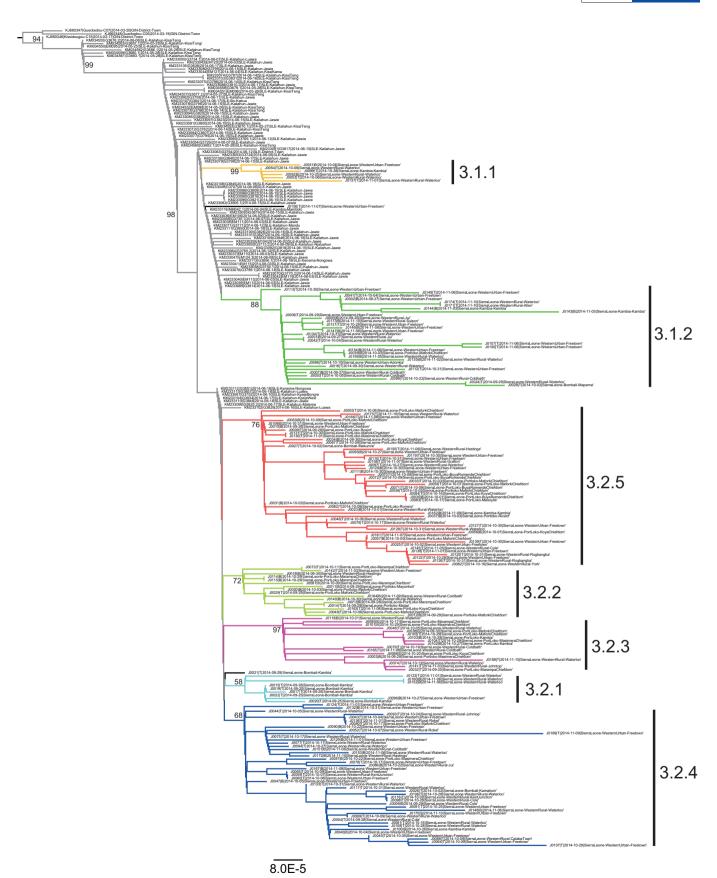
**Molecular characterizations of the 2014 EBOV.** SNPs were called directly from the sequence alignment using the CLC Genomic Workbench v7.5.1, GeneiousR8 and Newbler v2.9. The earliest strain of EBOV 2014, H.sapiens-wt/GIN/2014/Makona-Kissidougou-C15 (GenBank accession number KJ660346.2) was used as the reference genome. The synonymous substitutions, non-synonymous substitutions, and substitutions in non-coding regions were marked with coloured dots.

15. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
16. Bielejec, F. *et al.* SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics* **27**, 2910–2912 (2011).
17. Minin, V.N. *et al.* Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol. Biol. Evol.* **25**, 1459–1471 (2008).
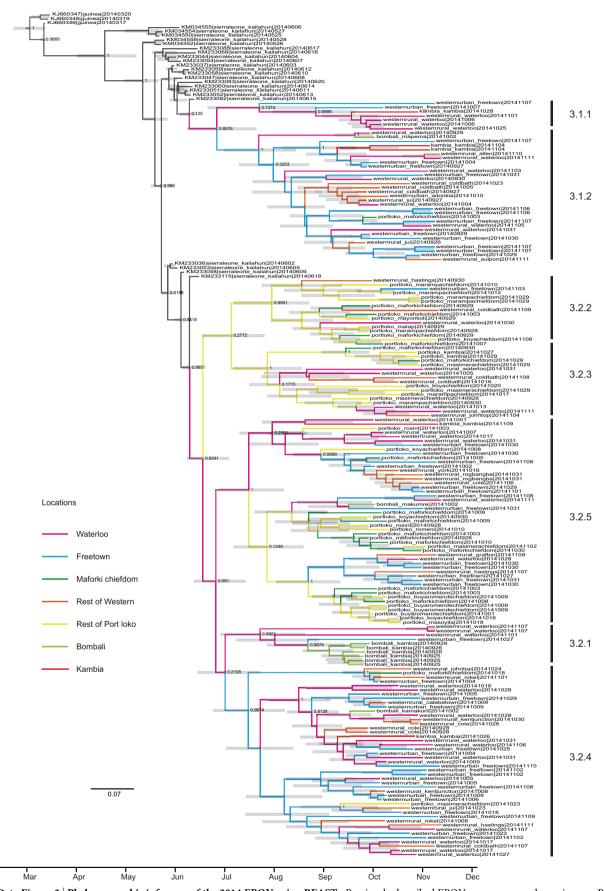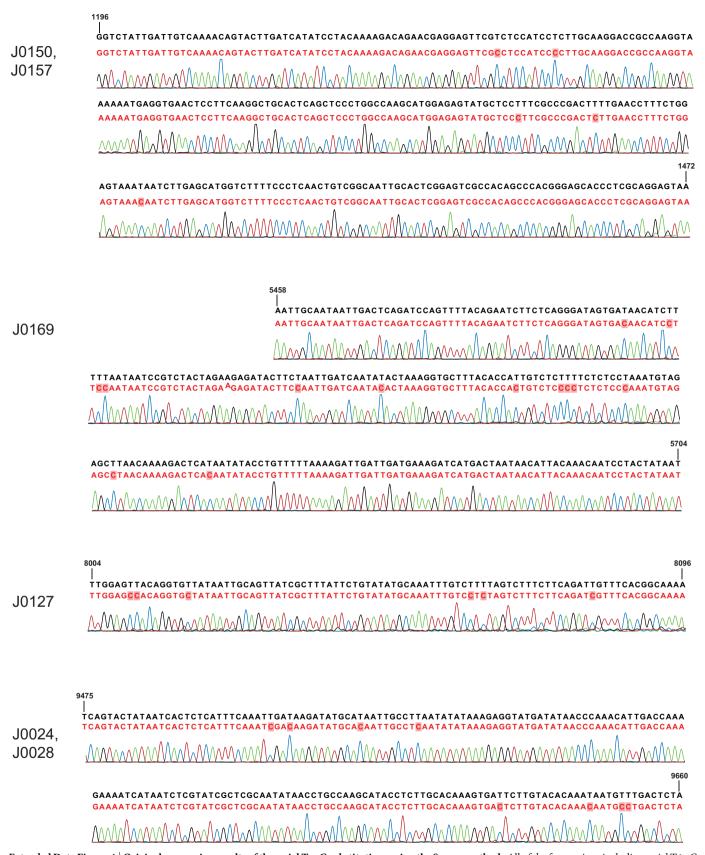
**Extended Data Figure 1 | Phylogenetic tree of the 2014 EBOV inferred using MrBayes.** The seven novel sublineages are highlighted using different colours. Previously described EBOV sequences are shown in grey. Posterior probability for each lineage is shown.

**Extended Data Figure 2 | Maximum likelihood tree of 2014 EBOV constructed using RAxML.**

**Extended Data Figure 3 | Phylogeographic inference of the 2014 EBOV using BEAST.** Previously described EBOV sequences are shown in grey. Posterior probability for each lineage is shown.

**1196**

J0150,
J0157

GGTCTATTGATTGTCAAAACAGTACTTGATCATATCCTACAAAAGACAGAACGAGGAGTTCGTCTCCATCCTCTTGCAAGGACCGCCAAGGTA
GGTCTATTGATTGTCAAAACAGTACTTGATCATATCCTACAAAAGACAGAACGAGGAGTTCG**C**CTCCATCC**C**CTTGCAAGGACCGCCAAGGTA

AAAAATGAGGTGAACTCCTTCAAGGCTGCACTCAGCTCCCTGGCCAAGCATGGAGAGTATGCTCCTTTCGCCCGACTTTTGAACCTTTCTGG
AAAAATGAGGTGAACTCCTTCAAGGCTGCACTCAGCTCCCTGGCCAAGCATGGAGAGTATGCTCC**C**TTCGCCCGACT**C**TTGAACCTTTCTGG

AGTAAATAATCTTGAGCATGGTCTTTTCCCTCAACTGTCGGCAATTGCACTCGGAGTCGCCACAGCCCACGGGAGCACCCTCGCAGGAGTAA
AGTAAA**C**AATCTTGAGCATGGTCTTTTCCCTCAACTGTCGGCAATTGCACTCGGAGTCGCCACAGCCCACGGGAGCACCCTCGCAGGAGTAA

**1472**

**5458**

J0169

AATTGCAATAATTGACTCAGATCCAGTTTTACAGAATCTTCTCAGGGATAGTGATAACATCTT
AATTGCAATAATTGACTCAGATCCAGTTTTACAGAATCTTCTCAGGGATAGTGA**C**AACATC**C**T

TTTAATAATCCGTCTACTAGAAGAGATACTTCTAATTGATCAATATACTAAAGGTGCTTTACACCATTGTCTCTTTTCTCTCCTAAATGTAG
TCCAATAATCCGTCTACTAGA^A^GAGATACTTC**C**AATTGATCAATA**C**ACTAAAGGTGCTTTACACCAC**C**TGTCTC**CC**TCTCTCCC**C**AAATGTAG

AGCTTAACAAAAGACTCATAATATACCTGTTTTTAAAAGATTGATTGATGAAAGATCATGACTAATAACATTACAAACAATCCTACTATAAT
AGCC**C**TAACAAAAGACTCA**C**AATATACCTGTTTTTAAAAGATTGATTGATGAAAGATCATGACTAATAACATTACAAACAATCCTACTATAAT

**5704**

**8004**                                                                                                                    **8096**

J0127

TTGGAGTTACAGGTGTTATAATTGCAGTTATCGCTTTATTCTGTATATGCAAATTTGTCTTTTAGTCTTTCTTCAGATTGTTTCACGGCAAAA
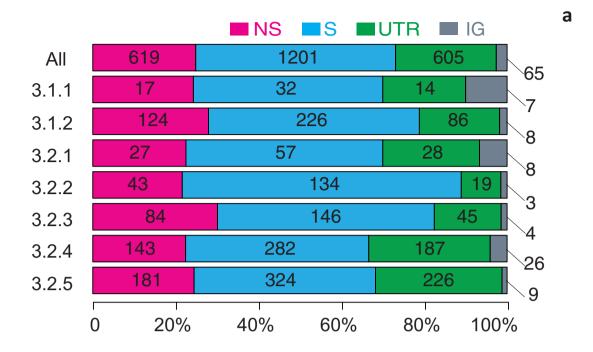TTGGAG**CC**ACAGGTG**C**TATAATTGCAGTTATCGCTTTATTCTGTATATGCAAATTTGTC**CT**CTAGTCTTTCTTCAGAT**C**GTTTCACGGCAAAA

**9475**

J0024,
J0028

TCAGTACTATAATCACTCTCATTTCAAATTGATAAGATATGCATAATTGCCTTAATATATAAAGAGGTATGATATAACCCAAACATTGACCAAA
TCAGTACTATAATCACTCTCATTTCAAAT**CGA**CAAGATATGCAC**A**ATTGCCT**C**AATATATAAAGAGGTATGATATAACCCAAACATTGACCAAA

GAAAATCATAATCTCGTATCGCTCGCAATATAACCTGCCAAGCATACCTCTTGCACAAAGTGATTCTTGTACACAAATAATGTTTGACTCTA
GAAAATCATAATCTCGTATCGCTCGCAATATAACCTGCCAAGCATACCTCTTGCACAAAGTGA**C**TCTTGTACACAA**C**AATG**CC**TGACTCTA

**9660**

**Extended Data Figure 4 | Original sequencing results of the serial T > C substitutions using the Sanger method.** All of the four regions including serial T > C substitutions were sequenced using the Sanger method with the primers provided in Extended Data Table 1.

a



**Gene-specific dN/dS**

b

| Gene | dN/dS | 95% HPD |
|------|-------|---------|
| NP | 0.4069 | (0.2429, 0.6324) |
| VP35 | 0.4497 | (0.2409, 0.7548) |
| VP40 | 0.2339 | (0.0839, 0.5029) |
| GP | 0.6065 | (0.3828, 0.9047) |
| VP30 | 0.2359 | (0.0938, 0.4782) |
| VP24 | 0.1114 | (0.0064, 0.4907) |
| L | 0.2463 | (0.1777, 0.3308) |

**Lineage-specific dN/dS**

c

| Lineage | dN/dS | 95% HPD |
|---------|-------|---------|
| 3.1.1 | 0.5847 | (0.2096, 1.2562) |
| 3.1.2 | 0.2769 | (0.1593, 0.4415) |
| 3.2.1 | 0.4242 | (0.1822, 0.8202) |
| 3.2.2 | 0.2087 | (0.1003, 0.3762) |
| 3.2.3 | 0.3869 | (0.2179, 0.6263) |
| 3.2.4 | 0.2879 | (0.1837, 0.4255) |
| 3.2.5 | 0.3414 | (0.2301, 0.4839) |

**Extended Data Figure 5 | Synonymous and non-synonymous substitutions of the 2014 EBOV. a**, Distribution of synonymous and non-synonymous substitutions in different lineages. The numbers of substitutions are labelled within bars. NS, non-synonymous; S, synonymous; UTR, UTR region; IG, intergenic. **b**, Gene-specific global dN/dS estimates. The dN/dS and 95% highest posterior density interval were calculated using HyPhy. **c**, Lineage-specific global dN/dS estimates.

**Extended Data Table 1 | Primers designed for the confirmation of the serial T > C substitutions**

| Sample ID | Target region | Primer |
|---|---|---|
| J0150 & J0157 | 1139-1474 | GGACATGATGCCAACGATGC(+) |
| | | ATTTACTCCTGCGAGGGTGC(-) |
| J0169 | 5434-5731 | GTCTTCCAGCTGTGGTTGAGA(+) |
| | | AAGATTGACATTTGAATCACCGT(-) |
| J0127 | 7967-8200 | TGGTGGACAGGATGGAGACA(+) |
| | | GGCTATGTTTGAAGCTCCAGTG(-) |
| J0024 & J0028 | 9401-9783 | CCTTCTACTTGATCACAATACTCCG(+) |
| | | CCTCCTCCACAACTTGAAGCA(-) |

# CORRECTIONS & AMENDMENTS

## Erratum: Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Leone

Yi-Gang Tong, Wei-Feng Shi, Di Liu, Jun Qian, Long Liang, Xiao-Chen Bo, Jun Liu, Hong-Guang Ren, Hang Fan, Ming Ni, Yang Sun, Yuan Jin, Yue Teng, Zhen Li, David Kargbo, Foday Dafae, Alex Kanu, Cheng-Chao Chen, Zhi-Heng Lan, Hui Jiang, Yang Luo, Hui-Jun Lu, Xiao-Guang Zhang, Fan Yang, Yi Hu, Yu-Xi Cao, Yong-Qiang Deng, Hao-Xiang Su, Yu Sun, Wen-Sen Liu, Zhuang Wang, Cheng-Yu Wang, Zhao-Yang Bu, Zhen-Dong Guo, Liu-Bo Zhang, Wei-Min Nie, Chang-Qing Bai, Chun-Hua Sun, Xiao-Ping An, Pei-Song Xu, Xiang-Li-Lan Zhang, Yong Huang, Zhi-Qiang Mi, Dong Yu, Hong-Wu Yao, Yong Feng, Zhi-Ping Xia, Xue-Xing Zheng, Song-Tao Yang, Bing Lu, Jia-Fu Jiang, Brima Kargbo, Fu-Chu He, George F. Gao, Wu-Chun Cao
& The China Mobile Laboratory Testing Team in Sierra Leone

This Letter should have contained an associated Creative Commons statement in the Author Information section. In addition, the Fig. 3c legend should have stated that the bar chart was adapted, with permission, from Ebola response roadmap - Situation report, Figure 3; http://www.who.int/csr/disease/ebola/situation-reports/en (accessed 1 April 2015)[1]. These issues have now both been corrected in the online versions of the paper.

1. World Health Organization. Ebola response roadmap - Situation report. http://www.who.int/csr/disease/ebola/situation-reports/en (accessed 1 April 2015).