

# Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa

Miles W. Carroll<sup>1,2,3</sup>, David A. Matthews<sup>4\*</sup>, Julian A. Hiscox<sup>5\*</sup>, Michael J. Elmore<sup>1\*</sup>, Georgios Pollakis<sup>5\*</sup>, Andrew Rambaut<sup>6,7,8\*</sup>, Roger Hewson<sup>1,2,9</sup>, Isabel García-Dorival<sup>5</sup>, Joseph Akoi Bore<sup>2,10,11</sup>, Raymond Koundouno<sup>2,10,11</sup>, Saïd Abdellati<sup>2,12</sup>, Babak Afrough<sup>1,2</sup>, John Aiyepada<sup>2,13</sup>, Patience Akhilomen<sup>2,13</sup>, Danny Asogun<sup>2,13</sup>, Barry Atkinson<sup>1,2</sup>, Marlis Badusche<sup>2,14,15</sup>, Amadou Bah<sup>2,16</sup>, Simon Bate<sup>1,2</sup>, Jan Baumann<sup>2,14</sup>, Dirk Becker<sup>2,15,17</sup>, Beate Becker-Ziaja<sup>2,14,15</sup>, Anne Bocquin<sup>2,18,19</sup>, Benny Borremans<sup>2,20</sup>, Andrew Bosworth<sup>1,2,5</sup>, Jan Peter Boettcher<sup>2,21</sup>, Angela Cannas<sup>2,22</sup>, Fabrizio Carletti<sup>2,22</sup>, Concetta Castilletti<sup>2,22</sup>, Simon Clark<sup>1,2</sup>, Francesca Colavita<sup>2,22</sup>, Sandra Diederich<sup>2,15,23</sup>, Adomeh Donatus<sup>2,13</sup>, Sophie Duraffour<sup>2,14,24</sup>, Deborah Ehichioya<sup>2,14,25</sup>, Heinz Ellerbrok<sup>2,21</sup>, Maria Dolores Fernandez-Garcia<sup>2,26</sup>, Alexandra Fizez<sup>2,18,27</sup>, Erna Fleischmann<sup>2,15,28</sup>, Sophie Gryseels<sup>2,20</sup>, Antje Hermelink<sup>2,21</sup>, Julia Hinzmann<sup>2,21</sup>, Ute Hopf-Guevara<sup>2,21</sup>, Yemisi Ighodalo<sup>2,13</sup>, Lisa Jameson<sup>1,2</sup>, Anne Kelterbaum<sup>2,15,17</sup>, Zoltan Kis<sup>2,29</sup>, Stefan Kloth<sup>2,21</sup>, Claudia Kohl<sup>2,21</sup>, Miša Korva<sup>2,30</sup>, Annette Kraus<sup>2,31</sup>, Eeva Kuisma<sup>1,2</sup>, Andreas Kurth<sup>2,21</sup>, Britta Liedigk<sup>2,14,15</sup>, Christopher H. Logue<sup>1,2</sup>, Anja Lütke<sup>2,15,32</sup>, Piet Maes<sup>2,24</sup>, James McCowen<sup>1,2</sup>, Stéphane Mély<sup>2,18,19</sup>, Marc Mertens<sup>2,15,23</sup>, Silvia Meschi<sup>2,22</sup>, Benjamin Meyer<sup>2,15,33</sup>, Janine Michel<sup>2,21</sup>, Peter Molkenthin<sup>2,15,28</sup>, César Muñoz-Fontela<sup>2,15,32</sup>, Doreen Muth<sup>2,15,33</sup>, Edmund N. C. Newman<sup>1,2</sup>, Didier Ngabo<sup>1,2</sup>, Lisa Oestereich<sup>2,14,15</sup>, Jennifer Okosun<sup>2,13</sup>, Thomas Olorok<sup>2,13</sup>, Racheal Omiunu<sup>2,13</sup>, Emmanuel Omomoh<sup>2,13</sup>, Elisa Pallasch<sup>2,14,15</sup>, Bernadett Pályi<sup>2,29</sup>, Jasmine Portmann<sup>2,34</sup>, Thomas Pottage<sup>1,2</sup>, Catherine Pratt<sup>1,2</sup>, Simone Priesnitz<sup>2,35</sup>, Serena Quartu<sup>2,22</sup>, Julie Rappe<sup>2,36</sup>, Johanna Repits<sup>2,37</sup>, Martin Richter<sup>2,21</sup>, Martin Rudolf<sup>2,14,15</sup>, Andreas Sachse<sup>2,21</sup>, Kristina Maria Schmidt<sup>2,21</sup>, Gordian Schudt<sup>2,15,17</sup>, Thomas Strecker<sup>2,15,17</sup>, Ruth Thom<sup>1,2</sup>, Stephen Thomas<sup>1,2</sup>, Ekaete Tobin<sup>2,13</sup>, Howard Tolley<sup>1,2</sup>, Jochen Trautner<sup>2,38</sup>, Tine Vermoesen<sup>2,12</sup>, Inês Vitoriano<sup>1,2</sup>, Matthias Wagner<sup>2,15,28</sup>, Svenja Wolff<sup>2,15,17</sup>, Constanze Yue<sup>2,21</sup>, Maria Rosaria Capobianchi<sup>2,22</sup>, Birte Kretschmer<sup>39</sup>, Yper Hall<sup>1</sup>, John G. Kenny<sup>40</sup>, Natasha Y. Rickett<sup>5</sup>, Gytis Dudas<sup>6</sup>, Cordelia E. M. Coltart<sup>41</sup>, Romy Kerber<sup>2,14,15</sup>, Damien Steer<sup>42</sup>, Callum Wright<sup>43</sup>, Francis Senyah<sup>1</sup>, Sakoba Keita<sup>44</sup>, Patrick Drury<sup>45</sup>, Boubacar Diallo<sup>46</sup>, Hilde de Clerck<sup>47</sup>, Michel Van Herp<sup>47</sup>, Armand Sprecher<sup>47</sup>, Alexis Traore<sup>48</sup>, Mandiou Diakite<sup>49</sup>, Mandy Kader Konde<sup>50</sup>, Lamine Koivogui<sup>11</sup>, N'Faly Magassouba<sup>10</sup>, Tatjana Avšič-Županc<sup>2,30</sup>, Andreas Nitsche<sup>2,21</sup>, Marc Strasser<sup>2,34</sup>, Giuseppe Ippolito<sup>2,22</sup>, Stephan Becker<sup>2,15,17</sup>, Kilian Stoecker<sup>2,15,28</sup>, Martin Gabriel<sup>2,14,15</sup>, Hervé Raoul<sup>2,19</sup>, Antonino Di Caro<sup>2,22</sup>, Roman Wölfel<sup>2,15,28</sup>, Pierre Formenty<sup>45</sup> & Stephan Günther<sup>2,14,15\*</sup>

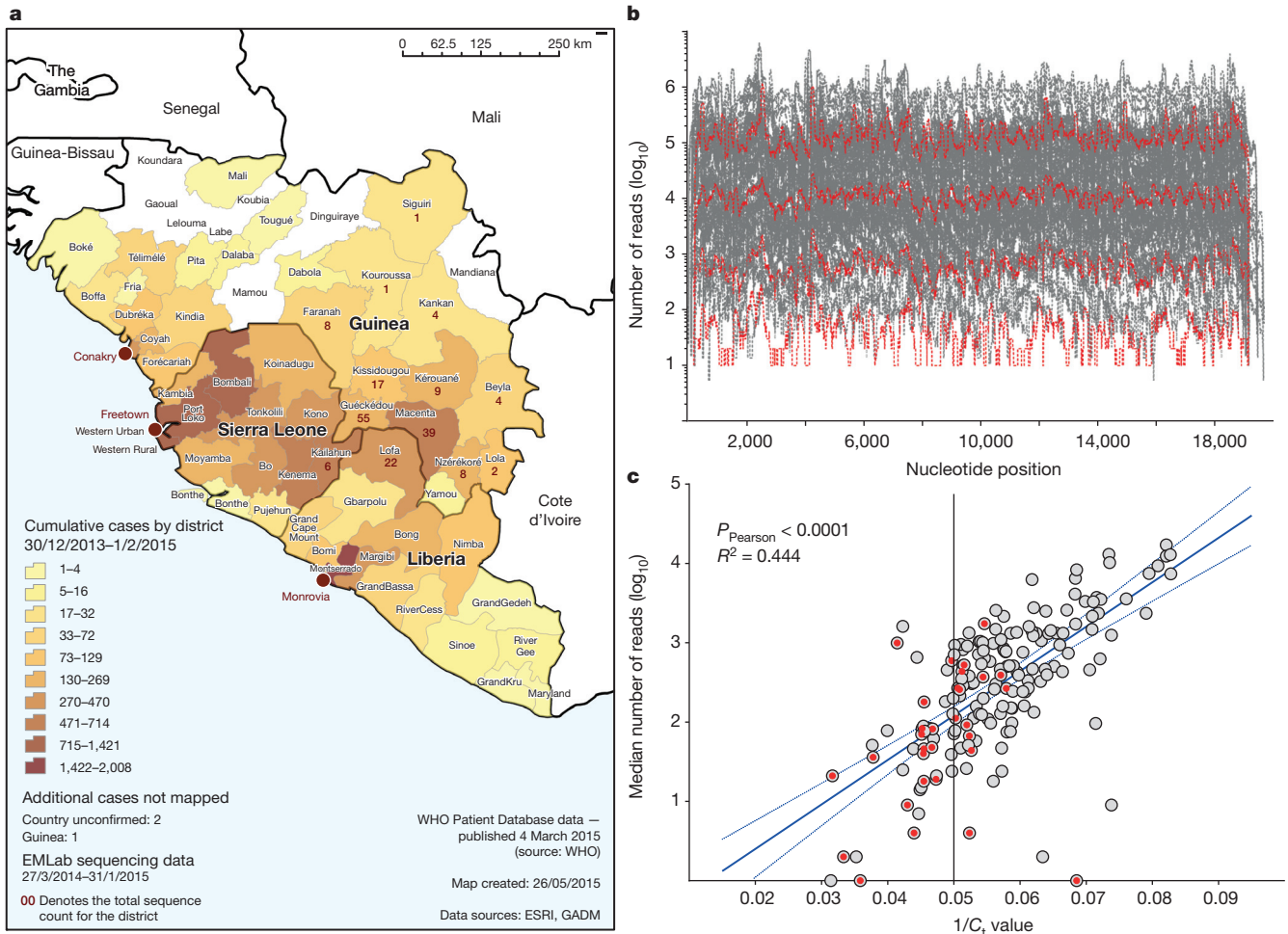
West Africa is currently witnessing the most extensive Ebola virus (EBOV) outbreak so far recorded<sup>1–3</sup>. Until now, there have been 27,013 reported cases and 11,134 deaths. The origin of the virus is thought to have been a zoonotic transmission from a bat to a two-year-old boy in December 2013 (ref. 2). From this index case the virus was spread by human-to-human contact throughout Guinea, Sierra Leone and Liberia. However, the origin of the particular virus in each country and time of transmission is not known and currently relies on epidemiological analysis, which may be unreliable owing to the difficulties of obtaining patient information. Here we trace the genetic evolution of EBOV in the current outbreak that has resulted in multiple lineages. Deep sequencing of 179 patient samples processed by the European Mobile Laboratory, the first diagnostics unit to be deployed to the epicentre of the outbreak in Guinea, reveals an epidemiological and evolutionary

history of the epidemic from March 2014 to January 2015. Analysis of EBOV genome evolution has also benefited from a similar sequencing effort of patient samples from Sierra Leone. Our results confirm that the EBOV from Guinea moved into Sierra Leone, most likely in April or early May. The viruses of the Guinea/Sierra Leone lineage mixed around June/July 2014. Viral sequences covering August, September and October 2014 indicate that this lineage evolved independently within Guinea. These data can be used in conjunction with epidemiological information to test retrospectively the effectiveness of control measures, and provides an unprecedented window into the evolution of an ongoing viral haemorrhagic fever outbreak.

We used a deep sequencing approach to gain insight into the evolution of Ebola virus (EBOV) in Guinea from the ongoing West African outbreak. This was an approach based on analysis pipelines developed

<sup>1</sup>Public Health England, Porton Down, Wiltshire SP4 0JG, UK. <sup>2</sup>The European Mobile Laboratory Consortium, Bernhard-Nocht-Institute for Tropical Medicine, D-20359 Hamburg, Germany. <sup>3</sup>University of Southampton, South General Hospital, Southampton SO16 6YD, UK. <sup>4</sup>Department of Cellular and Molecular Medicine, School of Medical Sciences, University of Bristol, Bristol BS8 1TD, UK. <sup>5</sup>Institute of Infection and Global Health, University of Liverpool, Liverpool L69 2BE, UK. <sup>6</sup>Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 2FL, UK. <sup>7</sup>Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>8</sup>Centre for Immunology, Infection and Evolution, University of Edinburgh, Edinburgh EH9 2FL, UK. <sup>9</sup>London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. <sup>10</sup>Université Gamal Abdel Nasser de Conakry, Laboratoire des Fièvres Hémorragiques en Guinée, Conakry, Guinea. <sup>11</sup>Institut National de Santé Publique, Conakry, Guinea. <sup>12</sup>Institute of Tropical Medicine, B-2000 Antwerp, Belgium. <sup>13</sup>Institute of Lassa Fever Research and Control, Irrua Specialist Teaching Hospital, Irrua, Edo State, Nigeria. <sup>14</sup>Bernhard Nocht Institute for Tropical Medicine, D-20359 Hamburg, Germany. <sup>15</sup>German Centre for Infection Research (DZIF), 38124 Braunschweig, Germany. <sup>16</sup>Swiss Tropical and Public Health Institute, University of Basel, CH-4002 Basel, Switzerland. <sup>17</sup>Institute of Virology, Philipps University Marburg, 35043 Marburg, Germany. <sup>18</sup>National Reference Center for Viral Hemorrhagic Fevers, 69365 Lyon, France. <sup>19</sup>Laboratoire P4 Inserm-Jean Mérioux, US003 Inserm, 69365 Lyon, France. <sup>20</sup>Department of Biology, University of Antwerp, B-2020 Antwerp, Belgium. <sup>21</sup>Robert Koch Institute, 13353 Berlin, Germany. <sup>22</sup>National Institute for Infectious Diseases (INMI) Lazzaro Spallanzani, 00149 Rome, Italy. <sup>23</sup>Friedrich Loeffler Institute, Federal Research Institute for Animal Health, 17493 Greifswald, Insel Riems, Germany. <sup>24</sup>KU Leuven Rega Institute, B-3000 Leuven, Belgium. <sup>25</sup>Redeemer's University, Osun State, Nigeria. <sup>26</sup>Centro Nacional de Microbiología, Instituto de Salud Carlos III, 28029 Madrid, Spain. <sup>27</sup>Unité de Biologie des Infections Virales Emergentes, Institut Pasteur, 69365 Lyon, France. <sup>28</sup>Bundeswehr Institute of Microbiology, 80937 Munich, Germany. <sup>29</sup>National Center for Epidemiology, National Biosafety Laboratory, H-1097 Budapest, Hungary. <sup>30</sup>Institute of Microbiology and Immunology, Faculty of Medicine, University of Ljubljana, SI-1000 Ljubljana, Slovenia. <sup>31</sup>Public Health Agency of Sweden, 171 82 Solna, Sweden. <sup>32</sup>Heinrich Pette Institute – Leibniz Institute for Experimental Virology, 20251 Hamburg, Germany. <sup>33</sup>Institute of Virology, University of Bonn, 53127 Bonn, Germany. <sup>34</sup>Federal Office for Civil Protection, Spiez Laboratory, CH-3700 Spiez, Switzerland. <sup>35</sup>Bundeswehr Hospital, 22049 Hamburg, Germany. <sup>36</sup>Institute of Virology and Immunology, CH-3147 Mittelhäusern, Switzerland. <sup>37</sup>Janssen-Cilag, SE-192 07 Sollentuna, Sweden. <sup>38</sup>Thünen Institute, D-22767 Hamburg, Germany. <sup>39</sup>Eurice - European Research and Project Office GmbH, 10115 Berlin, Germany. <sup>40</sup>Centre for Genomic Research, Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK. <sup>41</sup>Department of Infection and Population Health, University College London, London WC1E 6JB, UK. <sup>42</sup>Research IT, University of Bristol, Bristol BS8 1HH, UK. <sup>43</sup>Advanced Computing Research Centre, University of Bristol, Bristol BS8 1HH, UK. <sup>44</sup>Ministry of Health Guinea, Conakry, Guinea. <sup>45</sup>World Health Organization, 1211 Geneva 27, Switzerland. <sup>46</sup>World Health Organization, Conakry, Guinea. <sup>47</sup>Médecins Sans Frontières, B-1050 Brussels, Belgium. <sup>48</sup>Section Prévention et Lutte contre la Maladie à la Direction Préfectorale de la Santé de Guéckédou, Guéckédou, Guinea. <sup>49</sup>Université Gamal Abdel Nasser de Conakry, CHU Donka, Conakry, Guinea. <sup>50</sup>Health and Sustainable Development Foundation, Conakry, Guinea.

\*These authors contributed equally to this work.



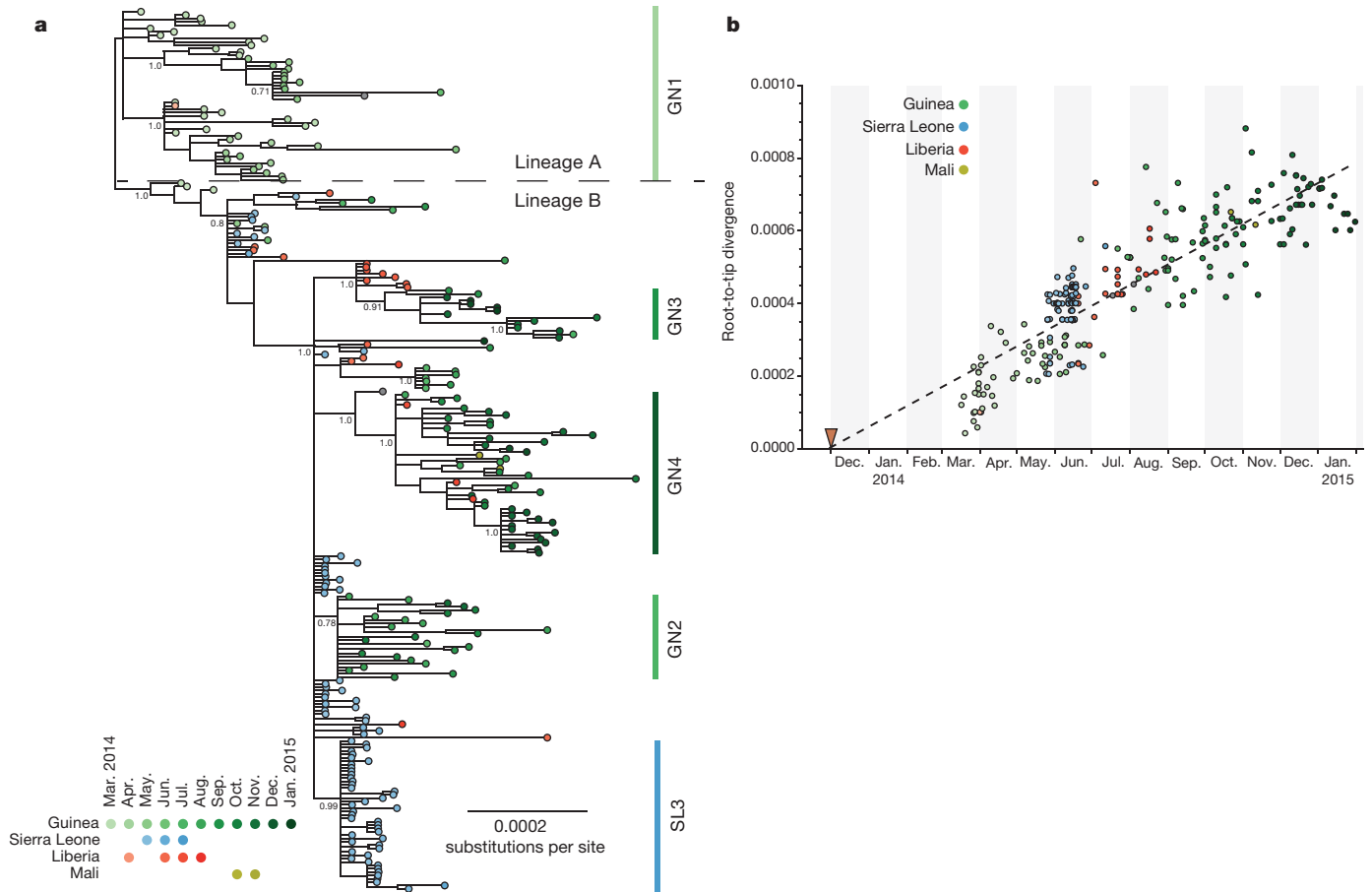
**Figure 1 | Geographical location, sequence read depth, and read depth vs  $C_t$  value of patient samples.** **a**, Geographical location of patient samples. The origin of the sequenced samples (one sample per patient) from Guinea, Sierra Leone, and Liberia processed by EMLab Guéckédou are plotted as numbers of cases by district. EMLab data are overlaid on an Ebola outbreak distribution map where cumulative cases are plotted as a heat map (low (yellow) to high (brown)) of confirmed cases from March 2014 to January 2015. Case data sourced from World Health Organization (WHO) Ebola response situation reports (<http://apps.who.int/ebola/en/ebola-situation-reports>); Geographic Information Systems (GIS) data sourced from Environmental Systems Research Institute (ESRI) and Database of Global Administrative Areas

for a guinea-pig model of EBOV infection and Hendra virus infection of human and bat cells<sup>4,5</sup>. Here we use this approach to derive consensus EBOV genomes from individual patient samples that can be used to study viral genome evolution during the course of the outbreak. Viral genomes were derived primarily from blood samples that had been taken from patients in Guinea and sent to the European Mobile Laboratory (EMLab), deployed by the World Health Organisation within the Médecins Sans Frontières Ebola Treatment Centre Guéckédou in March 2014 to aid the diagnostic effort. With the permission of Guinean authorities a biobank of samples was assembled which had known provenance of EBOV infection. Linked to each sample were the following data: patient location (to district level), sample collection date, disease onset and outcome. The collection dates were a median of 4 days after the date of onset of symptoms. Baseline data was cleaned, formatted and imported into the Geographic Information System, ESRI ArcGIS. Statistical tools were used to generate tabular output and to join the numeric case data with the district level boundaries of Guinea, Liberia and Sierra Leone (district geometries freely available from <http://www.gadm.org/>) (Fig. 1a).

(GADM; <http://www.gadm.org/>). **b**, Sequence depth per nucleotide position. The number of reads for each nucleotide position was plotted across the full length of the virus genome for each of the 179 virus isolates we analysed. In red is shown the uniformity of the depth across individual genomes, although the median number of reads per nucleotide position had a variation spanning over four log<sub>10</sub> units. **c**, Linear regression of the log<sub>10</sub> median sequence depth of each virus isolate versus the C<sub>t</sub> value of the viral load as determined by qRT-PCR. Red dots indicate samples taken from patients who went on to survive EBOV infection and grey shaded dots are from patients who records suggest died from EBOV infection.

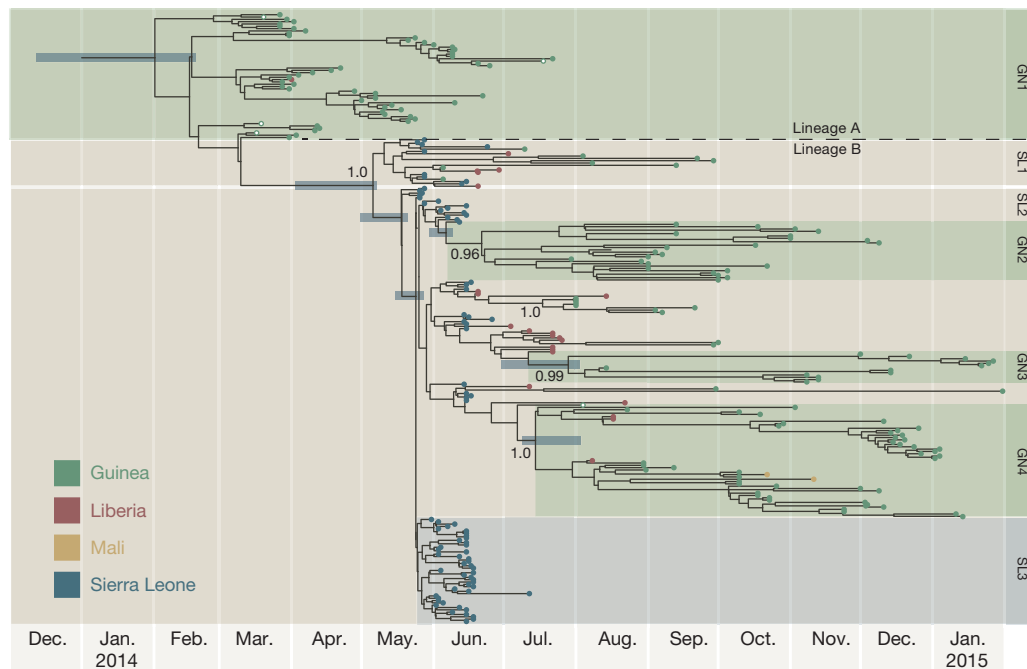
The viral genome sequence was derived from RNA sequencing analysis of the patient samples with no pre-amplification of the viral genome. In general we selected a range of samples from both males and females of different ages and a fair representation of sequences for each month (Extended Data Fig. 1), and with C<sub>t</sub> values less than 20 for EBOV RNA. In this selected patient cohort, with a relatively high viral load, there was approximately 80% mortality. The read depth mapping to the EBOV genome varied between samples and regions in the genome (Fig. 1b) and in general the number of sequence reads obtained for each genome correlated with the amount of viral load as determined by quantitative reverse-transcription PCR (qRT-PCR) (Fig. 1c).

Phylogenetic analysis revealed the dynamic nature of the epidemic and molecular change in the viral sequence (Fig. 2a). Several distinct lineages were identified, with an initial lineage A (Figs 2a, 3 and Extended Data Fig. 2) linked to early Guinean cases dating from March 2014 including the three original viruses published by Baize *et al.*<sup>2</sup>. A second lineage, B, emerged in May and June and comprises all the sequences from Gire *et al.*<sup>6</sup> and the remainder of those described

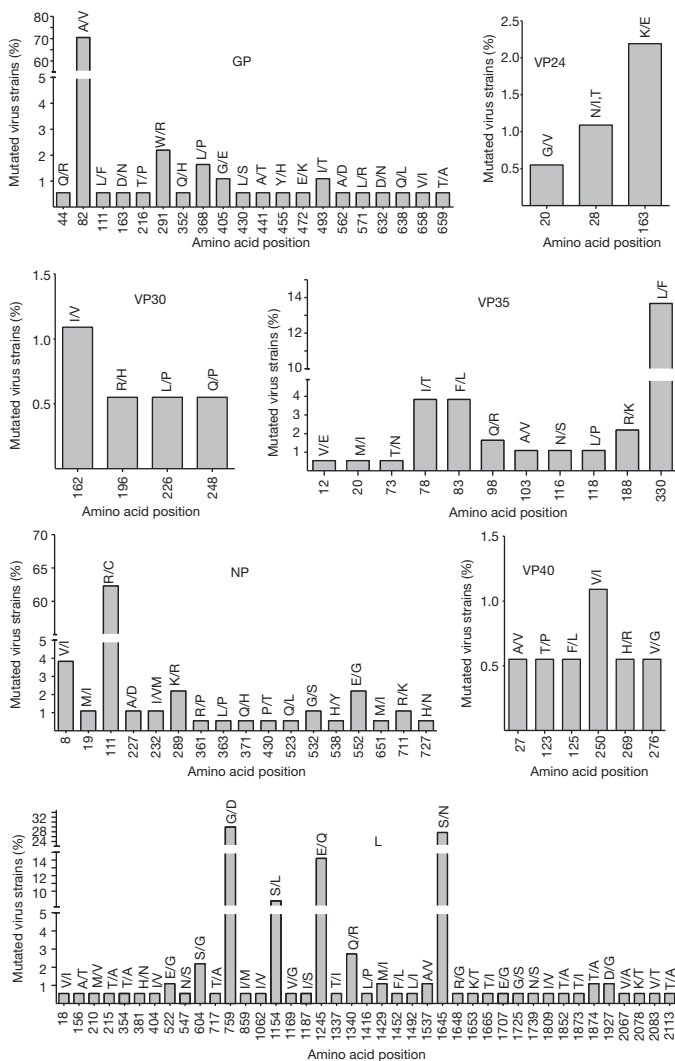


**Figure 2 | Phylogenetic relatedness and nucleotide sequence divergence of EBOV isolates from the 2013–2015 outbreak.** **a**, Phylogenetic relatedness of EBOV isolates. Phylogenetic tree inferred using MrBayes<sup>11</sup> for full-length EBOV genomes sequenced from 179 patient samples obtained between March 2014 and January 2015. Displayed is the majority consensus of 10,000 trees sampled from the posterior distribution with mean branch lengths. Posterior support is shown for selected key nodes. Twenty-two samples originated in Liberia and were collected between March and August 2014 and six samples

from Sierra Leone were obtained in June and July 2014. In our analysis we also included published sequences, including the three early Guinean sequences<sup>2</sup> and 78 sequences described by Gire *et al.*<sup>6</sup>. A number of lineages predominantly circulating in Guinea are denoted as GN1–4 along with a uniquely Sierra Leone lineage (SL3) recognised in Gire *et al.*<sup>6</sup>. **b**, EBOV nucleotide sequence divergence from root of the phylogeny in Fig. 2a plotted against time of collection of each virus. The date of the first documented case near Meliandou in eastern Guinea is indicated by the red triangle.



**Figure 3 | A time-scaled phylogenetic tree of 262 EBOV genomes from Guinea, Sierra Leone, Liberia and Mali.** Shown is a maximum clade credibility tree constructed from 10,000 trees sampled from the posterior distribution with mean node ages. Clades described in Gire *et al.*<sup>6</sup> are identified here (SL1, SL2 and SL3) as well as a number of lineages predominantly circulating in Guinea and posterior probability support is given for these. For certain key node ages, 95% credible intervals are shown by horizontal bars.



**Figure 4** | Position of non-synonymous amino acid variations in the 179 genomes analysed in this study compared to a reference sequence taken from March 2014 (KJ660346.2). Shown is the frequency of all amino acid positions that had variability and the substitution that occurred with the first single letter position indicating the reference sequence and the second position showing the variation. The percentage frequency in the 179 genomes is shown on the y axis. GP, glycoprotein; NP, nucleoprotein; L, RNA polymerase; VP, viral protein.

here. As the epidemic expanded, lineage A remained confined in Guinea from March to June 2014, except for one sequence from 18 July 2014. A single Liberian sequence from March 2014 grouped within this lineage. No further EBOV genomes that we sequenced from samples taken after July 2014 belonged to lineage A. This clade was likely to have been associated with the original outbreak in Guinea and was almost successfully contained in May 2014 by the interventions of the multi-agency response. Two clusters of Sierra Leone viruses described by Gire *et al.*<sup>6</sup> (denoted by the authors as clusters SL1 and SL2), both of which contain later viruses from Guinea and Liberia, suggest continued spread across the border during this time. Early cases in SL1 and SL2 were both associated with a single funeral<sup>6</sup>, so it is possible that this event may have reignited the epidemic. Thereafter, lineage B spread into Guinea, Liberia and Sierra Leone. This lineage is associated with the large epidemics in these three countries and persisted into 2015. The spatiotemporal spread of these viruses based on the phylogenetic analysis presented in Figs 2a and 3 was summarized (Extended Data Fig. 3) and indicated how the virus may have spread between the neighbouring countries. There was no

evidence from the data that increases or decreases in mortality were associated with any particular virus cluster (Extended Data Fig. 4).

The Bayesian time-scaled phylogenetic analysis estimated an average rate of evolution over the genome of  $1.42 \times 10^{-3}$  substitutions per site per year with 95% credible intervals of  $1.22 \times 10^{-3}$  and  $1.62 \times 10^{-3}$ . Details of the model assumptions are given in the Methods section. This rate is lower than that initially described for the West African outbreak by Gire *et al.*<sup>6</sup> but still higher than the long-term, between-outbreak rate of  $0.8 \times 10^{-3}$  estimated using viruses back to the 1976 Yambuku outbreak<sup>6</sup>. This apparent drop in rate of evolution between these two studies is consistent with the explanation provided by Gire *et al.*<sup>6</sup> that the short sampling interval (March to June) provided insufficient time for the action of purifying selection. However, the much longer sampling interval in the present study may simply be providing a more precise estimate of the rate. It should be noted, however, that the between-outbreak rate will exclusively reflect transmission and evolution that has occurred in the non-human reservoir species, so may not be directly comparable to the rate within a human outbreak. We observed no evidence of a change in evolutionary rate over the course of the epidemic with the accumulation of genetic change having a linear relationship with time (Fig. 2b), confirming that the apparent decline in rate between the two studies is an observational phenomenon<sup>7</sup> rather than a change in the virus.

The estimate of the date of the most recent common ancestor of the sampled viruses is mid-January 2014 (95% credible intervals 12 December 2013, 18 February 2014). Although this is an estimate of first transmission event that resulted in more than one lineage in our sample, this provides an upper bound on the date of emergence of the virus into the human population. This date estimate is consistent with the epidemiological tracing of the first suspected cases to December 2013<sup>2</sup>.

Given the error-prone nature of EBOV genome replication we examined the potential amino acid variation in EBOV proteins from the start of our sample collection in March 2014 to January 2015. The location of amino acid changes on EBOV proteins and their relative representation in the 179 assembled genomes were compared to an isolate identified in March 2014 (ref. 2) (Fig. 4). While there is amino acid variation in all of the genomes sampled, there were very few changes in viral protein 30 (VP30), viral protein 40 (VP40) and viral protein 24 (VP24), and these changes are only in less than ~2% of the genomes sampled. However, a single amino acid substitution in VP24 is associated with adaptation to a new host<sup>4,8</sup>, and this may be due to interactions with host-cell proteins<sup>9,10</sup>. While some of the variation may be attributed to a purely random molecular clock pattern, in GP, VP35, NP and L there are some amino acid variations that are present in over ~15% of the genomes sampled. For example, in GP there is an A to V substitution in ~70.5% of the genomes sampled compared to the reference genome. Implications of the mutations within GP in relation to immune escape of therapeutics and vaccines will need to be assessed in pseudotype neutralization assays using EBOV monoclonal antibodies and serum from people who have been vaccinated.

**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.

Received 9 April; accepted 1 June 2015.

Published online 17 June; corrected online 5 August 2015 (see full-text HTML version for details).

- Schieffelin, J. S. *et al.* Clinical illness and outcomes in patients with Ebola in Sierra Leone. *N. Engl. J. Med.* **371**, 2092–2100 (2014).
- Baize, S. *et al.* Emergence of Zaire Ebola virus disease in Guinea. *N. Engl. J. Med.* **371**, 1418–1425 (2014).
- Gatherer, D. The unprecedented scale of the West African Ebola virus disease outbreak is due to environmental and sociological factors, not special attributes of the currently circulating strain of the virus. *Evid. Based Med.* **20**, 28 (2015).

4. Dowall, S. D. *et al.* Elucidating variations in the nucleotide sequence of Ebola virus associated with increasing pathogenicity. *Genome Biol.* **15**, 540 (2014).
5. Wynne, J. W. *et al.* Proteomics informed by transcriptomics reveals Hendra virus sensitizes bat cells to TRAIL-mediated apoptosis. *Genome Biol.* **15**, 532 (2014).
6. Gire, S. K. *et al.* Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345**, 1369–1372 (2014).
7. Ho, S. Y., Phillips, M. J., Cooper, A. & Drummond, A. J. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol. Biol. Evol.* **22**, 1561–1568 (2005).
8. Mateo, M. *et al.* VP24 is a molecular determinant of Ebola virus virulence in guinea pigs. *J. Infect. Dis.* **204** (Suppl 3), S1011–S1020 (2011).
9. García-Dorival, I. *et al.* Elucidation of the Ebola virus VP24 cellular interactome and disruption of virus biology through targeted inhibition of host-cell protein function. *J. Proteome Res.* **13**, 5120–5135 (2014).
10. Basler, C. F. & Amarasinghe, G. K. Evasion of interferon responses by Ebola and Marburg viruses. *J. Interferon Cytokine Res.* **29**, 511–520 (2009).
11. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The authors would like to acknowledge that the EMLab response and the subsequent EBOV genome sequencing study would not have been possible without the extensive support from the many different agencies and organisations working in the West African EBOV disease outbreak region. EMLab worked with WHO, MSF and the Guinean authorities to tackle the outbreak in the Guéckédou area where the samples from this study were collected. We thank those who helped make this possible and the Guinean authorities for their decision to release the diagnostic samples to EMLab for shipment to Europe to undergo further analysis, including sequencing. We acknowledge Air France, Brussels Airlines and Virgin Airlines for transporting EMLab personnel and equipment in and out of West Africa during the outbreak period; World Courier for shipping our EBOV-positive samples out of Guinea to Europe; and the logistics support units and pilots and drivers of WHO/United Nations in West Africa for transporting our people and equipment throughout the region, and especially the drivers who made the 28 h round trip journey from Conakry to enable the EMLab unit to be established and resupplied in Guéckédou. We appreciate the work of the numerous European Embassies operating in West Africa who provided emergency support to our personnel at times of need. We thank M. Bull, J. Lewis, P. Payne and S. Leach from the Microbial Risk Assessment and Behavioural Science Team, Emergency Response Department, Public Health England; J. Tree from Public Health

England for help with GenBank submission; and S. Price and I. Stewart for helping with the running of our software on BlueCrystal, University of Bristol. We thank the people of West Africa for their gratitude and optimism, and for their positive attitude to our presence that we encountered on the daily journey to the Ebola Treatment Centre in Guéckédou. We acknowledge the efforts of the late Dr Lamine Ouendeno, who was one of the first healthcare workers to die during the current EBVD outbreak. We also thank Isabel and Maurice Ouendeno for providing us with food and shelter whilst delivering our Ebola response duties. This work was carried out in the context of the project EVIDENT (Ebola virus disease: correlates of protection, determinants of outcome, and clinical management) that received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 666100 and in the context of service contract IFS/2011/272-372 funded by Directorate-General for International Cooperation and Development. The EMLab is a technical partner in the WHO Emerging and Dangerous Pathogens Laboratory Network (EDPLN), and the Global Outbreak Alert and Response Network (GOARN) and the deployments in West Africa have been coordinated and supported by the GOARN Operational Support Team at WHO/HQ.

**Author Contributions** M.W.C., S.G., J.A.H., D.A.M. and N.M. designed the study. J.A.H., D.A.M., M.J.E., A.R., G.P., S.G. and M.W.C. wrote the manuscript. D.A.M., J.A.H., M.J.E., A.R., G.P., M.W.C., S.G., Y.H. and I.G.D. analysed the data. All other authors were involved either in sample collection, processing and/or logistical support and strategic oversight for the work.

**Author Information** The 179 consensus genome sequences described in this study have been assigned the GenBank accession numbers KR817067–KR817245. Further information is provided in Supplementary Table 1. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.W.C. ([miles.carroll@phe.gov.uk](mailto:miles.carroll@phe.gov.uk)).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

## METHODS

No statistical methods were used to predetermine sample size. There was no randomization or blinding in selection of samples for sequencing.

**Ethics statement.** The National Committee of Ethics in Medical Research of Guinea approved the use of diagnostic leftover samples and corresponding patient data for this study (permit no. 11/CNERS/14). As the samples had been collected as part of the public health response to contain the outbreak in Guinea, informed consent was not obtained from patients.

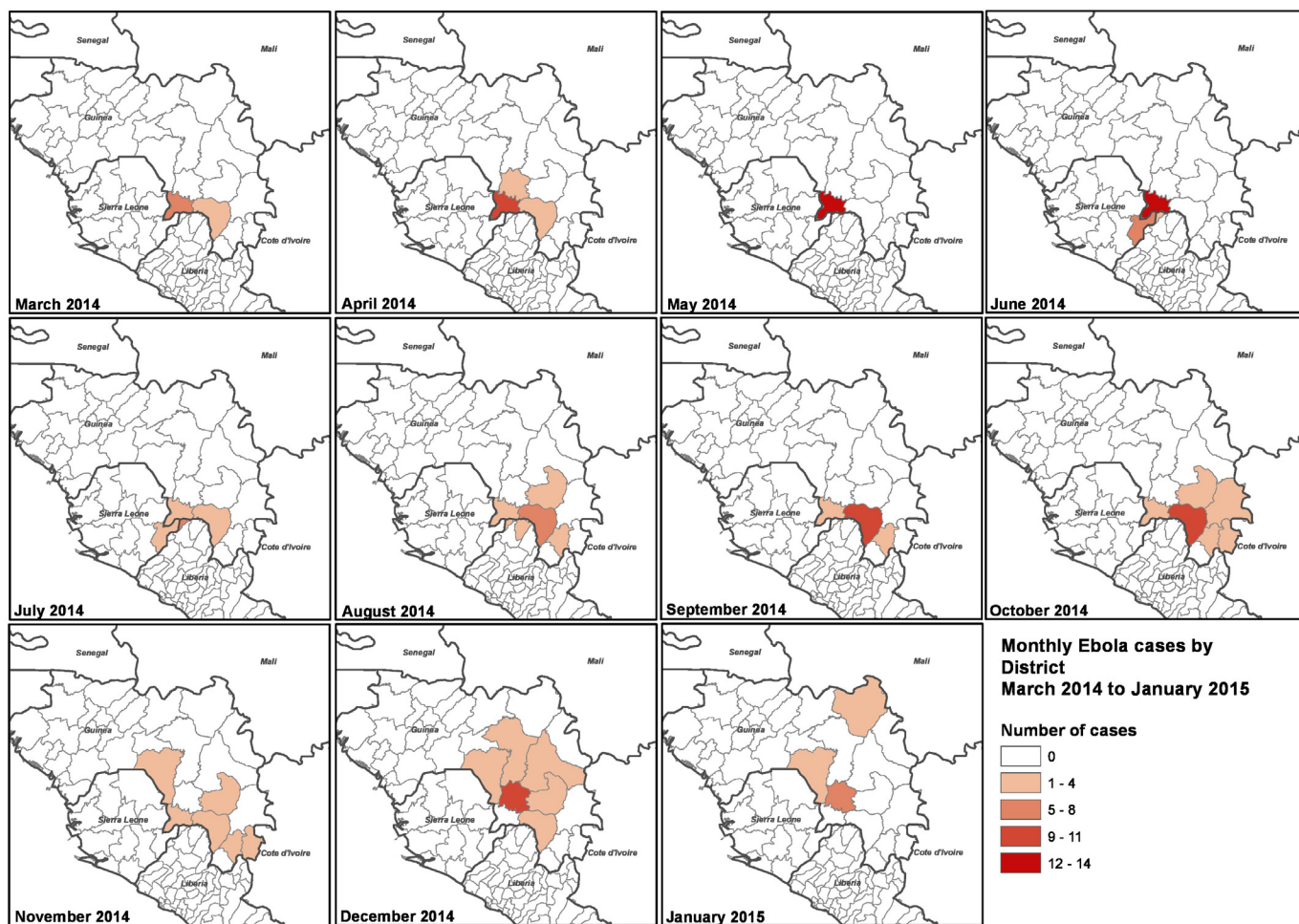
**Genome sequencing and consensus building.** Viral genome sequence was derived from the RNA extracted for diagnostic purposes from blood samples in the field with no pre-amplification of the viral genome. These samples were processed by the EMLab and are detailed in Supplementary Table 1, which indicates sample name, geographical location, date of onset of symptoms, date sample was collected, and the  $C_t$  value of EBOV RNA at the date of test. The clinical status is also indicated as well as malaria co-infection where known. Extracted RNA was DNase treated with Turbo DNase (Ambion) using the rigorous protocol. RNA sequencing libraries were prepared from the resultant RNA using the Epicentre ScriptSeq v2 RNA-Seq Library Preparation Kit. Following 10–15 cycles of amplification, libraries were purified using AMPure XP beads. Each library was quantified using Qubit and the size distribution assessed using the Agilent 2100 Bioanalyzer. These final libraries were pooled in equimolar amounts using the Qubit and Bioanalyzer data with 9–10 libraries per pool. The quantity and quality of the pool was assessed by Bioanalyzer and subsequently by qPCR using the Illumina Library Quantification Kit from Kapa on a Roche Light Cycler LC480II according to manufacturer's instructions. Each pool of libraries was sequenced on one lane of a HiSeq2500 at  $2 \times 125$ -bp paired-end sequencing with v4 chemistry.

The trimmed fastq files were first aligned to a copy of the human genome using Bowtie2 (ref. 12) and the unaligned reads were then mapped with Bowtie2 to a list of 3731 known viral genomes excluding EBOV genomes. The reads that were still unmapped were then aligned to the EBOV genome—either the prototype strain isolated in Zaire in 1976 (AF086833.2) or a strain isolated during the current outbreak (KJ660348.2). For this step we again used Bowtie2 and the resultant alignment files were filtered with samtools to remove unmapped reads and reads with a mapping quality score below 11, followed by filtering with markdup to remove PCR duplicates. The resultant BAM file was then analysed by Quasirecomb<sup>13</sup> to generate a phred-weighted table of nucleotide frequencies which were parsed with a custom perl script to generate a consensus genome in fasta format. This consensus genome was then used as a reference genome to which we remapped the sequence reads which did not map to the human genome or other

viruses in order to generate a second consensus. In this way we were able to manually determine if the reference genome used by Bowtie2 influenced the process of calling a consensus genome. In addition, we used FreeBayes to independently call and identify SNPs and indels. The pipeline is entirely open source and implemented in the Galaxy environment<sup>14</sup>, a Galaxy compatible workflow, novel scripts and XML wrappers needed for implementation in Galaxy are freely available and included in Supplementary Data File 1. Sequence alignment maps were manually inspected and curated over regions with consistent low coverage (for example, at the 5' ends).

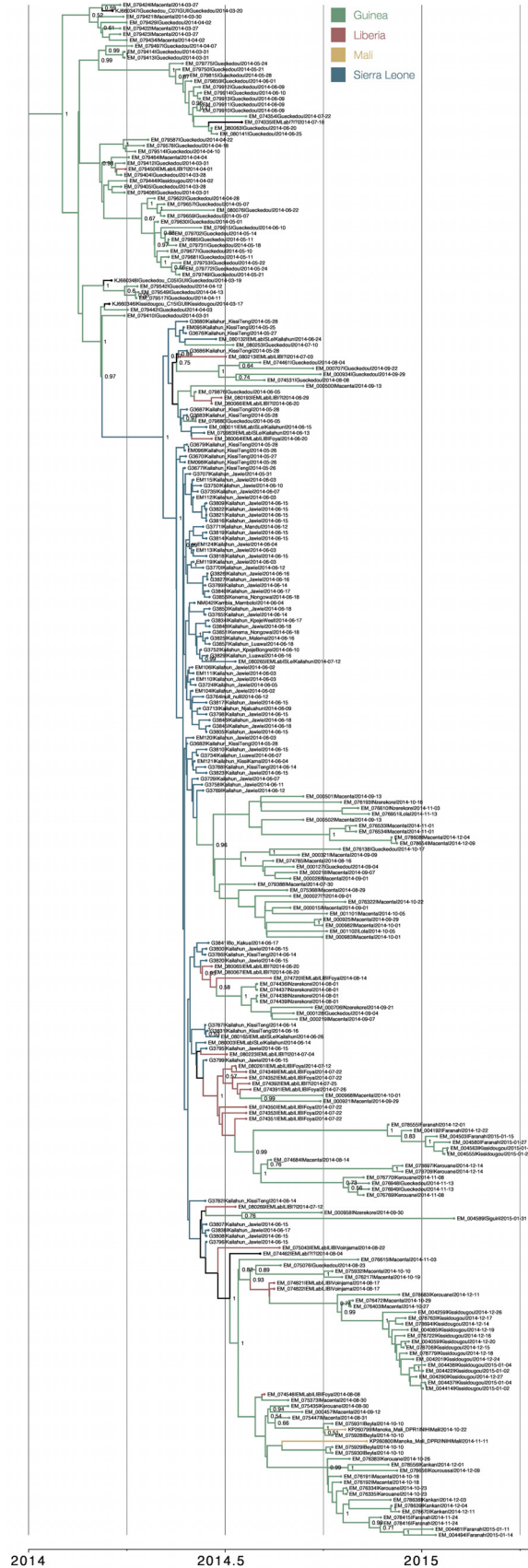
**Phylogenetic analysis.** Phylogenetic analysis comprised the 179 EBOV genomes from this study, 78 genomes from Sierra Leone<sup>6</sup>, three sequences from Guinea<sup>7</sup> and two sampled from Mali<sup>15</sup>. The genomes were partitioned into four sets of sites—1st, 2nd and 3rd codon positions of the protein-coding regions and the non-coding intergenic regions—with each partition being assigned a generalized time reversible substitution model<sup>16</sup>, gamma distributed rate heterogeneity<sup>17</sup> and a relative rate of evolution. This model was used to construct a Bayesian nucleotide divergence tree (Fig. 2) using MrBayes<sup>11</sup> and a time-scaled phylogenetic analysis (Fig. 3) using BEAST<sup>18</sup> with a log-normal distributed relaxed molecular clock<sup>19</sup>, and the 'Skygrid' non-parametric coalescent tree prior<sup>20</sup>. The alignments and control files for both analyses are available in Supplementary Data Files 2 and 3 and provide documentation of all model parameters.

12. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).
13. Topfer, A. *et al.* Probabilistic inference of viral quasispecies subject to recombination. *J. Comput. Biol.* **20**, 113–123 (2013).
14. Goecks, J., Nekrutenko, A., Taylor, J. & Galaxy, T. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **11**, R86 (2010).
15. Hoenen, T. *et al.* Mutation rate and genotype variation of Ebola virus from Mali case sequences. *Science*, (2015).
16. Tavaré, S. Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences in *Lectures on Mathematics in the Life Sciences* Vol. 17 (ed. Muir, R. M.) *Some Mathematical Questions in Biology: DNA Sequence Analysis* (American Mathematical Society, 1986).
17. Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39**, 306–314 (1994).
18. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
19. Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).
20. Gill, M. S. *et al.* Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30**, 713–724 (2013).



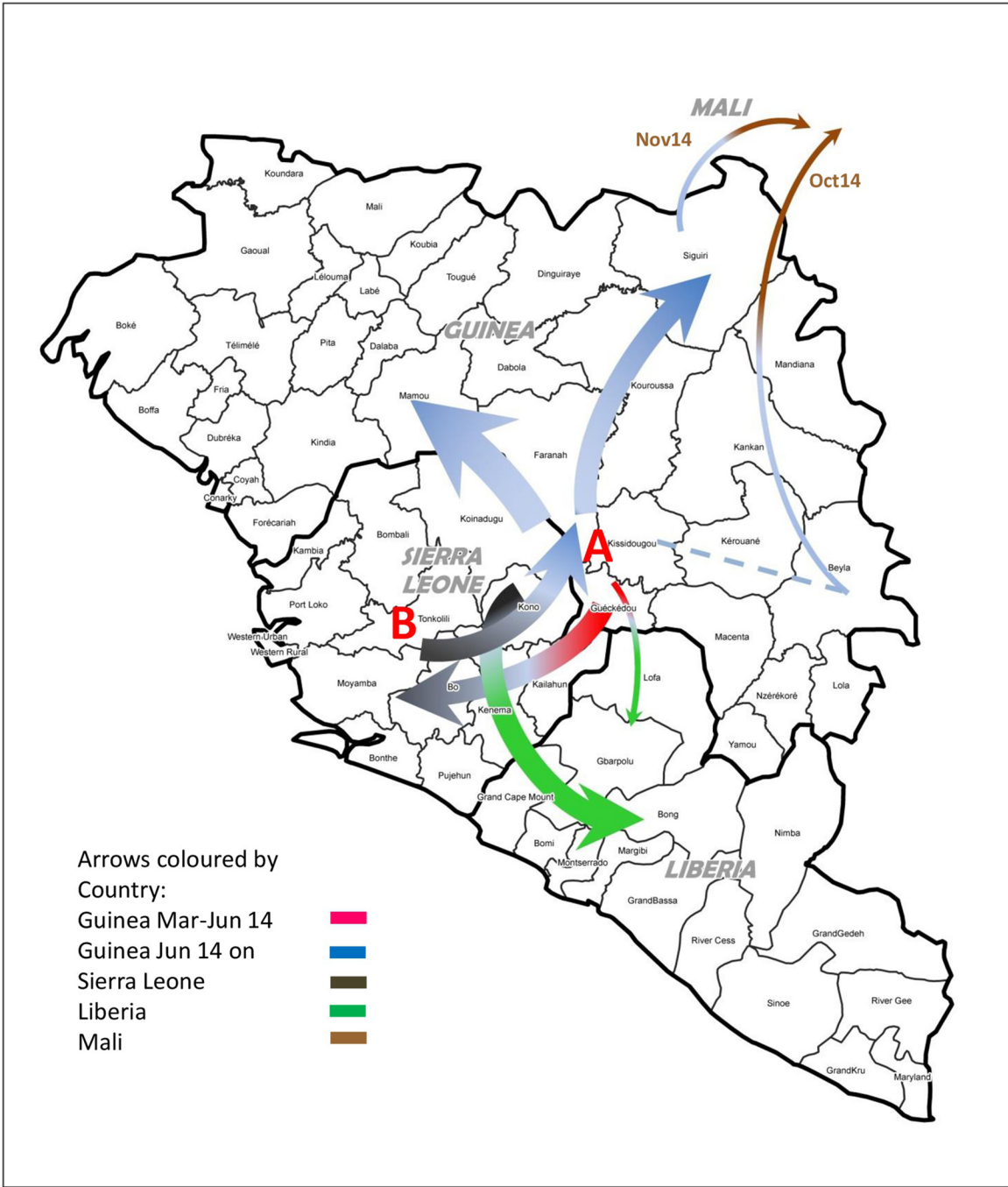
**Extended Data Figure 1 | Spatial and temporal location of patient samples. Geographical locations of sequenced samples are plotted by district as panels for each month of collection (March 2014–January 2015).** In brief, the number of samples obtained for each month was as follows: March 2014, 11;

April 2014, 14; May 2014, 14; June 2014, 22; July 2014, 16; August 2014, 19; September 2014, 18; October 2014, 21; November 2014, 11; December 2014, 22; January 2015, 11. Total number of samples sequenced, 179.



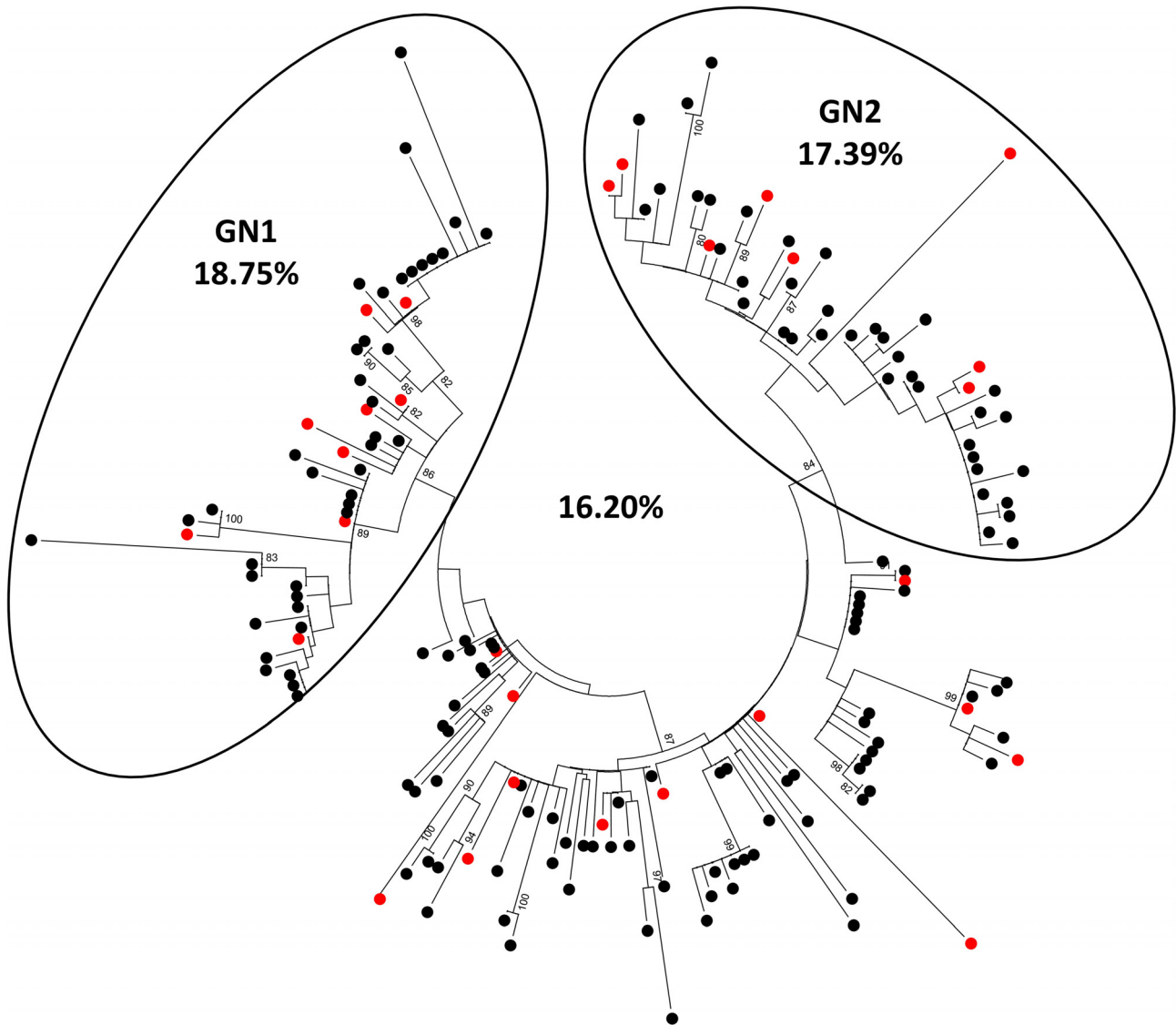
Extended Data Figure 2 | Enlarged view of phylogenetic tree presented in Fig. 3. Posterior support shown where >0.5.





**Extended Data Figure 3 | Temporal spread of EBOV based on phylogenetic analyses in Figs 2a and 3.** Colour scheme is as follows: Guinea is red/blue (1st half/2nd half of 2014, respectively), Sierra Leone is grey-black, Liberia is green, Mali is brown. Lineage A (A) is associated with the initial focus of the outbreak (Guéckédou, Macenta and Kissidougou) in March 2014, expanded around this area and then declined around July 2014. From lineage A a second lineage (B) emerged in May/June 2014 and expanded into Sierra Leone (end of May 2014) and Liberia (small arrow). Lineage B continued to spread into Sierra Leone, Liberia, and further into Guinea (beyond the original focus into most districts of Guinea). EBOV disease entered Mali from Guinea via two separate routes (from the Beyla district (possibly originally from Kissidougou) in October 2014 and from the Siguiri district in November 2014).

emerged in May/June 2014 and expanded into Sierra Leone (end of May 2014) and Liberia (small arrow). Lineage B continued to spread into Sierra Leone, Liberia, and further into Guinea (beyond the original focus into most districts of Guinea). EBOV disease entered Mali from Guinea via two separate routes (from the Beyla district (possibly originally from Kissidougou) in October 2014 and from the Siguiri district in November 2014).



**Extended Data Figure 4 | Survival rate amongst individuals with known EBOV sequences.** The total survival rate for the 179 sequenced virus isolates included in this study is presented, as is the survival rate for two sub-lineages, GN1 and GN2, as defined by phylogenetic inference in Figs 2a and 3. The

sequences available for GN1 were collected during the period of March–July 2014 and the sequences available for GN2 were collected during the period of August 2014–January 2015. Red dots indicate survivors.