LETTER

doi:10.1038/nature16996

Real-time, portable genome sequencing for Ebola surveillance

Joshua Quick[1]*, Nicholas J. Loman[1]*, Sophie Duraffour[2,3]*, Jared T. Simpson[4,5]*, Ettore Severi[6]*, Lauren Cowley[7]*, Joseph Akoi Bore[2], Raymond Koundouno[2], Gytis Dudas[8], Amy Mikhail[7], Nobila Ouédraogo[9], Babak Afrough[2,10], Amadou Bah[2,11], Jonathan H. J. Baum[2,3], Beate Becker-Ziaja[2,3], Jan Peter Boettcher[2,12], Mar Cabeza-Cabrerizo[2,3], Álvaro Camino-Sánchez[2], Lisa L. Carter[2,13], Juliane Doerrbecker[2,3], Theresa Enkirch[2,14], Isabel García-Dorival[2,15], Nicole Hetzelt[2,12], Julia Hinzmann[2,12], Tobias Holm[2,3], Liana Eleni Kafetzopoulou[2,16], Michel Koropogui[2,17], Abigael Kosgey[2,18], Eeva Kuisma[2,10], Christopher H. Logue[2,10], Antonio Mazzarelli[2,19], Sarah Meisel[2,3], Marc Mertens[2,20], Janine Michel[2,12], Didier Ngabo[2,10], Katja Nitzsche[2,3], Elisa Pallasch[2,3], Livia Victoria Patrono[2,3], Jasmine Portmann[2,21], Johanna Gabriella Repits[2,22], Natasha Y. Rickett[2,15,23], Andreas Sachse[2,12], Katrin Singethan[2,24], Inês Vitoriano[2,10], Rahel L. Yemanaberhan[2,3], Elsa G. Zekeng[2,15,23], Trina Racine[25], Alexander Bello[25], Amadou Alpha Sall[26], Ousmane Faye[26], Oumar Faye[26], N'Faly Magassouba[27], Cecelia V. Williams[28,29], Victoria Amburgey[28,29], Linda Winona[28,29], Emily Davis[29,30], Jon Gerlach[29,30], Frank Washington[29,30], Vanessa Monteil[31], Marine Jourdain[31], Marion Bererd[31], Alimou Camara[31], Hermann Somlare[31], Abdoulaye Camara[31], Marianne Gerard[31], Guillaume Bado[31], Bernard Baillet[31], Déborah Delaune[32,33], Koumpingnin Yacouba Nebie[34], Abdoulaye Diarra[34], Yacouba Savane[34], Raymond Bernard Pallawo[34], Giovanna Jaramillo Gutierrez[35], Natacha Milhano[6,36], Isabelle Roger[34], Christopher J. Williams[6,37], Facinet Yattara[17], Kuiama Lewandowski[10], James Taylor[38], Phillip Rachwal[38], Daniel J. Turner[39], Georgios Pollakis[15,23], Julian A. Hiscox[15,23], David A. Matthews[40], Matthew K. O'Shea[41], Andrew McD. Johnston[41], Duncan Wilson[41], Emma Hutley[42], Erasmus Smit[43], Antonino Di Caro[2,19], Roman Wölfel[2,44], Kilian Stoecker[2,44], Erna Fleischmann[2,44], Martin Gabriel[2,3], Simon A. Weller[38], Lamine Koivogui[45], Boubacar Diallo[34], Sakoba Keïta[17], Andrew Rambaut[8,46,47], Pierre Formenty[34], Stephan Günther[2,3] & Miles W. Carroll[2,10,48,49]

The Ebola virus disease epidemic in West Africa is the largest on record, responsible for over 28,599 cases and more than 11,299 deaths[1]. Genome sequencing in viral outbreaks is desirable to characterize the infectious agent and determine its evolutionary rate. Genome sequencing also allows the identification of signatures of host adaptation, identification and monitoring of diagnostic targets, and characterization of responses to vaccines and treatments. The Ebola virus (EBOV) genome substitution rate in the Makona strain has been estimated at between 0.87×10^{-3} and 1.42×10^{-3} mutations per site per year. This is equivalent to 16–27 mutations in each genome, meaning that sequences diverge rapidly enough to identify distinct sub-lineages during a prolonged epidemic[2–7]. Genome sequencing provides a high-resolution view of pathogen evolution and is increasingly sought after for outbreak surveillance. Sequence data may be used to guide control measures, but only if the results are generated quickly enough to inform interventions[8]. Genomic surveillance during the epidemic has been sporadic owing to a lack of local sequencing capacity coupled with practical difficulties transporting samples to remote sequencing facilities[9]. To address this problem, here we devise a genomic surveillance system that utilizes a novel nanopore DNA sequencing instrument. In April 2015 this system was transported in standard airline luggage to Guinea and used for real-time genomic surveillance of the ongoing epidemic. We present sequence data and analysis of 142 EBOV samples collected during the period March to October 2015. We were able to generate results less than 24 h after receiving an Ebola-positive sample, with the sequencing process taking as little as 15–60 min. We show that real-time genomic surveillance is possible in resource-limited settings and can be established rapidly to monitor outbreaks.

Conventional sequencing technologies are difficult to deploy in developing countries, where availability of continuous power and cold chains, laboratory space, and trained personnel is restricted. In addition, some genome sequencer instruments, such as those using optical

[1]Institute of Microbiology and Infection, University of Birmingham, Birmingham B15 2TT, UK. [2]The European Mobile Laboratory Consortium, Bernhard-Nocht-Institute for Tropical Medicine, D-20359 Hamburg, Germany. [3]Bernhard-Nocht-Institute for Tropical Medicine, D-20359 Hamburg, Germany. [4]Ontario Institute for Cancer Research, Toronto M5G 0A3, Canada. [5]Department of Computer Science, University of Toronto, Toronto M5S 3G4, Canada. [6]European Centre for Disease Prevention and Control (ECDC), 171 65 Solna, Sweden. [7]National Infection Service, Public Health England, London NW9 5EQ, UK. [8]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 2FL, UK. [9]Postgraduate Training for Applied Epidemiology (PAE, German FETP), Robert Koch Institute, D-13302 Berlin, Germany. [10]National Infection Service, Public Health England, Porton Down, Wiltshire SP4 0JG, UK. [11]Swiss Tropical and Public Health Institute, 4002 Basel, Switzerland. [12]Robert Koch Institute, D-13302 Berlin, Germany. [13]University College London, London WC1E 6BT, UK. [14]Paul-Ehrlich-Institut, Division of Veterinary Medicine, D-63225 Langen, Germany. [15]Institute of Infection and Global Health, University of Liverpool, Liverpool L69 7BE, UK. [16]Laboratory for Clinical and Epidemiological Virology, Department of Microbiology and Immunology, KU Leuven, Leuven B-3000, Belgium. [17]Ministry of Health Guinea, Conakry BP 585, Guinea. [18]Kenya Medical Research Institute, Nairobi P.O. BOX 54840 - 00200, Kenya. [19]National Institute for Infectious Diseases L. Spallanzani, 00149 Rome, Italy. [20]Friedrich-Loeffler-Institute, D-17493 Greifswald, Germany. [21]Federal Office for Civil Protection, Spiez Laboratory, 3700 Spiez, Switzerland. [22]Janssen-Cilag, Stockholm, Box 7073 – 19207, Sweden. [23]NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, University of Liverpool, Liverpool L69 7BE, UK. [24]Institute of Virology, Technische Universität München, D-81675 Munich, Germany. [25]Public Health Agency of Canada, Winnipeg, Manitoba R3E 3R2, Canada. [26]Institut Pasteur Dakar, Dakar, DP 220 Senegal. [27]Laboratoire de Fièvres Hémorragiques de Guinée, Conakry BP 5680, Guinea. [28]Sandia National Laboratories, PO Box 5800 MS1363, Albuquerque, New Mexico 87185-1363, USA. [29]Ratoma Ebola Diagnostic Center, Conakry, Guinea. [30]MRIGlobal, Kansas City, Missouri 64110-2241, USA. [31]Expertise France, Laboratoire K-plan de Forecariah en Guinée, 75006 Paris, France. [32]Fédération des Laboratoires - HIA Bégin, 94163 Saint-Mandé cedex, France. [33]Laboratoire de Biologie - Centre de Traitement des Soignants, Conakry, Guinea. [34]World Health Organization, Conakry BP 817, Guinea. [35]London School of Hygiene and Tropical Medicine, London EC1E 7HT, UK. [36]Norwegian Institute of Public Health, PO Box 4404 Nydalen, 0403 Oslo, Norway. [37]Public Health Wales, Cardiff CF11 9LJ, UK. [38]Defence Science and Technology Laboratory (Dstl) Porton Down, Salisbury SP4 0JQ, UK. [39]Oxford Nanopore Technologies, Oxford OX4 4GA, UK. [40]Department of Cellular and Molecular Medicine, School of Medical Sciences, University of Bristol, Bristol BS8 1TD, UK. [41]Academic Department of Military Medicine, Royal Centre for Defence Medicine, Birmingham B15 2TH, UK. [42]Centre of Defence Pathology, Royal Centre for Defence Medicine, Birmingham B15 2TH, UK. [43]Queen Elizabeth Hospital, Birmingham B12 2TH, UK. [44]Bundeswehr Institute of Microbiology, D-80937 Munich, Germany. [45]Institut National de Santé Publique, Conakry BP 1147, Guinea. [46]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892-2220, USA. [47]Centre for Immunology, Infection and Evolution, University of Edinburgh, Edinburgh EH9 2FL, UK. [48]University of Southampton, South General Hospital, Southampton SO16 6YD, UK. [49]NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, PHE Porton Down, UK.
*These authors contributed equally to this work.

© 2016 Macmillan Publishers Limited. All rights reserved

Figure 1 | Deployment of the portable genome surveillance system in Guinea. a, We were able to pack all instruments, reagents and disposable consumables within aircraft baggage. **b**, We initially established the genomic surveillance laboratory in Donka Hospital, Conakry, Guinea. **c**, Later we moved the laboratory to a dedicated sequencing laboratory in Coyah prefecture. **d**, Within this laboratory we separated the sequencing instruments (on the left) from the PCR bench (to the right). An uninterruptable power supply can be seen in the middle that provides power to the thermocycler. (Photographs taken by J.Q. and S.D.)

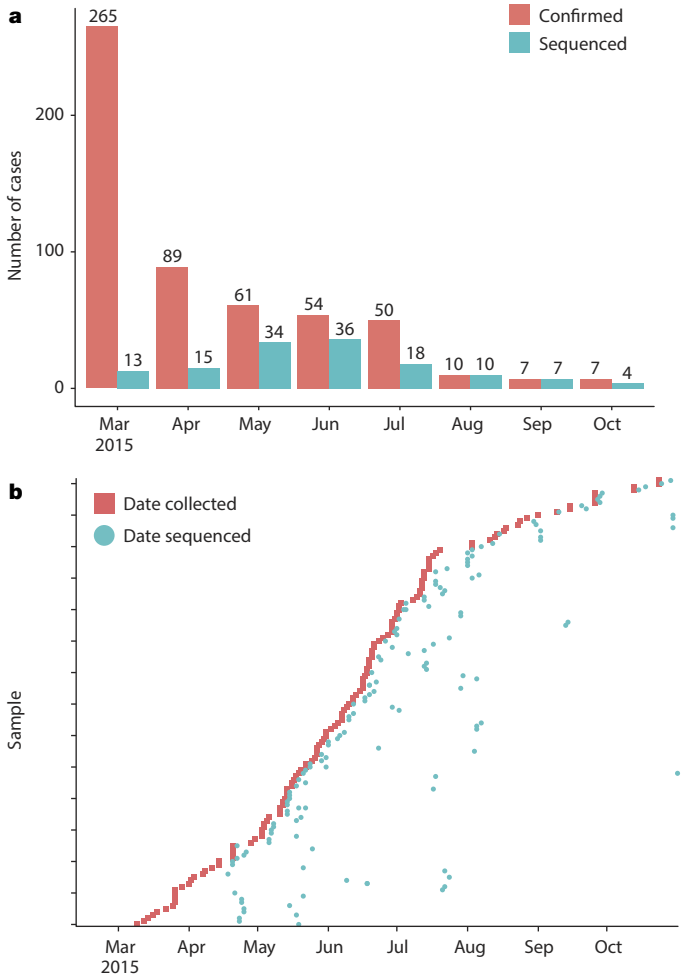


Figure 2 | Real-time genomics surveillance in context of the Guinea Ebola virus disease epidemic. a, Here we show the number of reported cases of Ebola virus disease in Guinea (red) in relation to the number of EBOV new patient samples ($n = 137$, in blue) generated during this study. **b**, For each of the 142 sequenced samples, we show the relationship between sample collection date (red) and the date of sequencing (blue). Twenty-eight samples were sequenced within three days of the sample being taken, and sixty-eight samples within a week. Larger gaps represent retrospective sequencing of cases to provide additional epidemiological context.

readings, for example the Illumina platform, require precise microscope alignment and repeated calibration by trained engineers[7,10]. Recently, a new highly portable genome sequencer has become available. The MinION (Oxford Nanopore Technologies, Oxford, UK) weighs less than 100 g. Data are read off the MinION from a laptop via a Universal Serial Bus (USB) port from which the instrument also draws power. The MinION works by taking frequent electrical current measurements as a single strand of DNA passes through a protein nanopore at 30 bases per second. DNA strands in the pore disrupts ionic flow, resulting in detectable changes in current that is dependent on the nucleotide sequence. Because the MinION detects single molecules it has a much higher error rate (between 10–20%[11,12]) than high-throughput instruments that read clonal copies of DNA molecules. Single-molecule sequencing has the advantage of being able to read extremely long molecules of DNA (50 kb or longer[12,13]). In order to generate accurate sequences, genomic regions must be read many times, with errors eliminated through consensus averaging. This system has previously been used to investigate a bacterial outbreak, but not yet a viral outbreak[14].

We designed a laboratory protocol to permit EBOV genome sequencing on the MinION that employed a targeted reverse transcriptase polymerase chain reaction (RT–PCR) in order to isolate sufficient DNA for sequencing. We considered and rejected an alternative approach, that of total RNA sequencing, as this approach also amplifies human-derived transcripts and dilutes viral signal[15]. We designed a panel of 38 primer pairs that would span the EBOV genome (Extended Data Fig. 1a, Supplementary Table 1). In pilot experiments at the Defence Science and Technology Laboratory (Dstl) Porton Down, UK, we sequenced a historic Zaire Ebolavirus using MinION as well as the Illumina MiSeq. Due to difficulties obtaining equal balancing of each of the 38 amplicon pairs only 65.7% of the EBOV genome was covered by at least 25 reads, compared with 87.4% on Illumina. However, nucleotide variants in those highly covered regions were concordant with those obtained from Illumina sequencing, with the exception of a single variant in a homopolymeric region. MinION sequencing currently cannot easily resolve the length of homopolymers of 5 bases or greater[16].

Next we designed a genome surveillance system that could be transported to West Africa. The system consisted of three MinION instruments (Oxford Nanopore Technologies, UK), four laptops, a thermocycler, a heat block, pipettes and sufficient reagents and consumables to perform sequencing (a full list of equipment is shown in Extended Data Fig. 2). We were able to pack this into less than 50 kg of standard airline travel luggage (Fig. 1a). We initially installed the genome surveillance system in the European Mobile Laboratory in Donka Hospital in Conakry, Guinea (Fig. 1b). Later on, the equipment was moved to a dedicated laboratory, located within the Coyah Ebola Treatment Unit (Fig. 1c, d).

We started sequencing genomes within 2 days of arriving in Guinea. We found early on that we were able to reliably generate long amplicons (around 2 kb in length) using primer pairs (Supplementary Table 4) in different combinations (Extended Data Fig. 1b, c). Using as few amplicons as possible significantly reduces effort when preparing samples. We found a combination of 11 amplicons that reliably amplified >97% of the EBOV genome.

We developed a bioinformatics approach that would yield accurate genotypes, and validated this using Makona virus samples from a previous study[3]. The bioinformatics workflow is detailed in the Methods and summarized in Extended Data Fig. 3. This validation process demonstrated that our bioinformatics analysis approach was robust. We compared our consensus sequences to those generated using Illumina

© 2016 Macmillan Publishers Limited. All rights reserved

sequencing and found that our approach was highly concordant, with no false positive variant calls. In several cases, we were unable to determine variants because they fell either within the primer binding region, or they were outside of the regions of the EBOV genome covered by our amplicon set (Extended Data Fig. 4a). These positions are represented as ambiguous nucleotides in the final consensus sequences used for analysis. Despite these masked positions, phylogenetic inference showed that samples clustered identically (Extended Data Fig. 4b). We determined that, despite the instrument's high error rate, use of electrical current information meant that 25-fold read coverage of genome positions was sufficient to determine accurate genotypes (Extended Data Fig. 5).

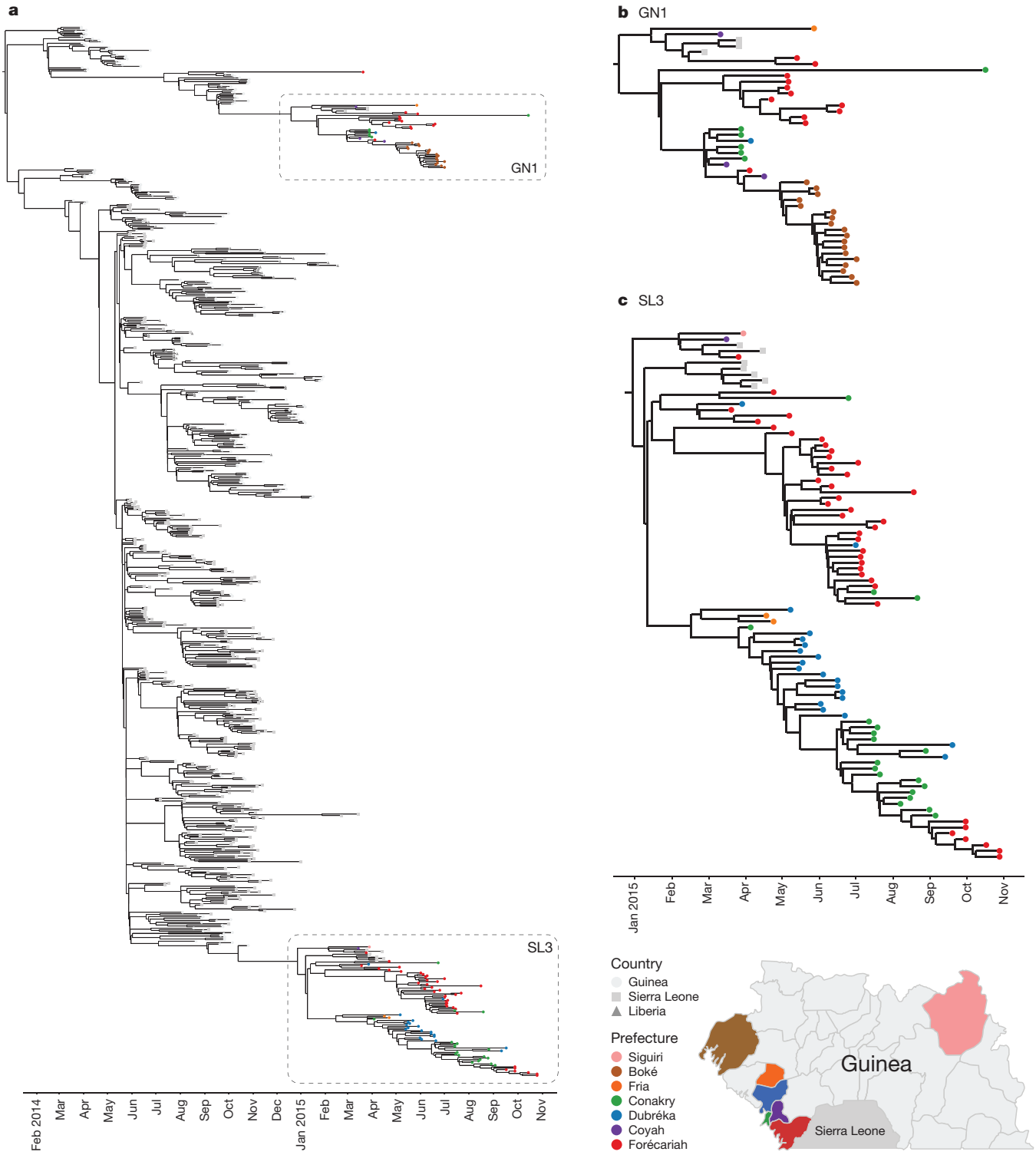


Figure 3 | Evolution of EBOV over the course of the Ebola virus disease epidemic. a, Time-scaled phylogeny of 603 published sequences with 125 high quality sequences from this study. The shape of nodes on the tree demonstrates country of origin. Our results show Guinean samples (coloured circles) belong to two previously identified lineages, GN1 and SL3. **b**, GN1 is deeply branching with early epidemic samples. **c**, SL3 is related to cases identified in Sierra Leone. Samples are frequently clustered by geography (indicated by colour of circle) and this provides information as to origins of new introductions, such as in the Boké epidemic in May 2015. Map figure adapted from SimpleMaps website (http://simplemaps.com/resources/svg-gn).

© 2016 Macmillan Publishers Limited. All rights reserved

After deployment of the genome surveillance system, we worked in partnership with diagnostic laboratories in Guinea to provide real-time sequencing results to National Coordination in Guinea and the World Health Organisation (WHO). Collaborating laboratories provided left-over diagnostic RNA extracts for sequencing. The genome sequencing workflow, including amplification, sequence library preparation and sequencing could be accomplished within a working day. In one case, including remote bioinformatics analysis, the fastest time from patient sample to answer was achieved in less than 24 h (Supplementary Table 1), although the protocol was more usually performed over two working days. We found that in half of cases, we were able to generate sufficient reads on the MinION (between around 5,000 and 10,000) in less than an hour (Extended Data Fig. 6). In total, 142 samples were sequenced over 148 MinION runs during the 6-month period, providing extensive coverage of reported cases in the outbreak (Fig. 2). Full details of samples and runs are in the Supplementary Data. We failed to generate amplicons for some samples, resulting in missing regions of the genome. Such samples often corresponded to those with a high RT–PCR cycle threshold (C_t) value, suggestive of lower viral loads (Extended Data Fig. 7). For these we used a modified RT–PCR scheme using 19 shorter amplicons. We assumed that difficulties generating long amplicons related to low numbers of starting molecules of that length in the original sample. We excluded 17 samples owing to quality control issues, for example single nucleotide polymorphism (SNP) calling sensitivity of less than 75%. We found that in-field performance of the system was comparable with validation experiments performed in the UK, suggesting that the system tolerated transportation well (Extended Data Fig. 8).

We combined our sequencing data set with 603 samples from other studies and inferred a time-scaled phylogenetic tree using the BEAST software package (Fig. 3). A maximum likelihood analysis and root-to-tip analysis showed good agreement between sampling date and root-to-tip divergence (Extended Data Figs 9 and 10a). We estimated a substitution rate of 1.19×10^{-3} (95% interval, 1.09×10^{-3}, 1.29×10^{-3}) of the combined data set (Extended Data Fig. 10b). This is consistent with rates from previous studies[2–7]. Results generated within the first 10 days of starting real-time sequencing indicated that the persisting Guinean cases belonged to two major lineages, named GN1 and SL3, that had been established near the beginning of the epidemic (Fig. 3). Lineage GN1 is deeply branching from early cases in Guinea and has been infrequently seen in Sierra Leone[2], suggesting that it has been largely confined to Guinea. The second lineage identified here was derived from lineage SL3 which was first detected in Sierra Leone by Gire *et al.*[2], but was later seen circulating in Conakry towards the end of 2014[3]. Through integration of our data set with those generated by a different group operating in Sierra Leone we detected that both GN1 and SL3 had also been seen in Sierra Leone early in 2015, suggestive of transmission between the countries[17].

This work demonstrates a step change in our ability to perform genomic surveillance prospectively during outbreaks under resource-limited conditions. However, numerous obstacles remain before such genomically informed investigations are routine. In practical terms, we encountered significant logistical issues when performing this work, notably the absence of reliable, continuous mains electrical power, forcing a dependence on unreliable electrical generators and uninterruptable power supply (UPS) units, particularly for the bulky PCR thermocyclers. However, portable, battery-powered thermocyclers are in development, and isothermal approaches may be preferable for future work[18]. By contrast, the MinION sequencer was unaffected by power outages and surges. We faced consistent issues with internet connectivity, which is currently required for analysis. There is a pressing need for a fully offline version of the analysis presented here. This would reduce the dependence on high bandwidth connections. However it is likely that phylogenetic analysis will continue to be performed remotely (discussed further in the supplementary Field Guide to Portable Sequencing). In this analysis we focused on variant calling approaches. A *de novo* approach to analysis would be preferable, but this would currently result in insertion and deletion errors due to poor resolution of homopolymeric tracts on the MinION. Our approach relies on amplification of genetic material before sequencing. In other epidemics, where the causative pathogen may be unidentified this is a drawback due to the need to have a priori knowledge of the pathogen genome sequence. In this event, sequencing directly from clinical material may be better, although sensitivity issues persist[15].

Real-time genomic surveillance is a new tool in our arsenal to assist difficult epidemiological investigations, and to provide an international and environmental context to emerging infectious diseases. This may improve the efficiency of resource allocation and the timeliness of epidemiological investigations through genomically informed investigations of transmission chains. Real-time genomic surveillance also increases the possibility of identifying previously unknown chains of transmission. By integrating in real time our data set with that of a second group performing sequencing in Sierra Leone, we identified evidence of frequent transmissions across the border with Guinea. Crucially, we released data at regular intervals throughout this project through Github, integrating our results with those of others, displayed interactively at http://ebola.nextstrain.org. We employed the Virological web forum to discuss complex cases (http://virological.org). This system will continue to support the West African epidemic response and will serve as a template for genomic surveillance of future outbreaks.

The Ebola epidemic was officially declared to be over on 14 January 2016 (http://www.who.int/mediacentre/news/releases/2016/ebola-zero-liberia/en/). Hours later, a new case of EVD was confirmed in Sierra Leone (http://who.int/mediacentre/news/statements/2016/new-ebola-case/en/), confirming warnings that further flare-ups may be expected. Such cases pose pressing questions about their source that may be answered through genomic surveillance, by determining links to previously infected individuals[19] and ruling out a new zoonotic spillover event. We now stand poised to answer such questions quickly.

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 18 November 2015; accepted 15 January 2016.
Published online 3 February 2016.

1. World Health Organisation. Ebola Situation Report - 11 November 2015. http://apps.who.int/ebola/current-situation/ebola-situation-report-11-november-2015 (2015).
2. Gire, S. K. *et al.* Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345,** 1369–1372 (2014).
3. Carroll, M. W. *et al.* Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. *Nature* **524,** 97–101 (2015).
4. Simon-Loriere, E. *et al.* Distinct lineages of Ebola virus in Guinea during the 2014 West African epidemic. *Nature* **524,** 102–104 (2015).
5. Park, D. J. *et al.* Ebola virus epidemiology, transmission, and evolution during seven months in Sierra Leone. *Cell* **161,** 1516–1526 (2015).
6. Tong, Y.-G. *et al.* Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Leone. *Nature* **524,** 93–96 (2015).
7. Kugelman, J. R. *et al.* Monitoring of Ebola Virus Makona evolution through establishment of advanced genomic capability in Liberia. *Emerg. Infect. Dis.* **21,** 1135–1143 (2015).
8. Gardy, J., Loman, N. J. & Rambaut, A. Real-time digital pathogen surveillance—the time is now. *Genome Biol.* **16,** 155 (2015).
9. Yozwiak, N. L., Schaffner, S. F. & Sabeti, P. C. Data sharing: make outbreak research open access. *Nature* **518,** 477–479 (2015).
10. Heger, M. Liberia's LIBR Genome Center monitors Ebola outbreak, emerging pathogens. https://www.genomeweb.com/sequencing-technology/liberias-libr-genome-center-monitors-ebola-outbreak-emerging-pathogens (2015).
11. Quick, J., Quinlan, A. R. & Loman, N. J. A reference bacterial genome dataset generated on the MinION™ portable single-molecule nanopore sequencer. *Gigascience* **3,** 22 (2014).
12. Jain, M. *et al.* Improved data analysis for the MinION nanopore sequencer. *Nature Methods* **12,** 351–356 (2015).
13. Urban, J. M., Bliss, J., Lawrence, C. E. & Gerbi, S. A. Sequencing ultra-long DNA molecules with the Oxford Nanopore MinION. Preprint at http://biorxiv.org/content/early/2015/05/13/019281 (2015).
14. Quick, J. *et al.* Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biol.* **16,** 114 (2015).

© 2016 Macmillan Publishers Limited. All rights reserved

15. Greninger, A. L. *et al.* Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Med.* **7,** 99 (2015).
16. Loman, N. J., Quick, J. & Simpson, J. T. A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nature Methods* **12,** 733–735 (2015).
17. Goodfellow, I. *et al.* Recent evolution patterns of Ebola virus obtained by direct sequencing in Sierra Leone. http://virological.org/t/recent-evolution-patterns-of-ebola-virus-obtained-by-direct-sequencing-in-sierra-leone/150 (2015).
18. Herold, K. E., Sergeev, N., Matviyenko, A. & Rasooly, A. in *Biosensors and Biodetection* **504,** 441–458 (Humana, 2009).
19. Mate, S. E. *et al.* Molecular evidence of sexual transmission of Ebola virus. *N. Engl. J. Med.* **373,** 2448–2454 (2015).

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

Acknowledgements 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 and the African Union. 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 program 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. J.Q. is funded by the NIHR Surgical Reconstruction and Microbiology Research Centre (SRMRC). N.J.L. is funded by a Medical Research Council Special Training Fellowship in Biomedical Informatics (to September 2015) and a Medical Research Council Bioinformatics Fellowship. J.T.S. is supported by the Ontario Institute for Cancer Research through funding provided by the Government of Ontario. Dstl support was funded by the UK Ministry of Defence (MOD). Dstl authors thank S. Lonsdale, C. Lonsdale and C. Mayers for supply of RNA, previous assistance, and review of the manuscript. The views expressed in this paper are not necessarily endorsed by the UK MOD. A.R. was supported by EU Seventh Framework Programme [FP7/2007-2013] under Grant Agreement no. 278433-PREDEMICS and ERC Grant agreement no. 260864. We are grateful for the generous support of University of Birmingham alumni for donations in support of the pilot work. The MRC Cloud Infrastructure for Microbial Bioinformatics (CLIMB) cyberinfrastructure was used to conduct bioinformatics analysis. The authors would like to thank B. Oppenheim and C. Wardius for help with logistics and the staff of Alta Biosciences, University of Birmingham and Sigma-Aldrich for generating PCR primers especially rapidly for this project. The authors would like to thank scientists deployed from the Special Pathogens Program from the National Microbiology Laboratory, Public Health Agency of Canada, who worked on EBOV diagnostics in Guinea. We are grateful to I. Goodfellow, M. Cotten and P. Kellam for permission to include sequences from Sierra Leone in this analysis. We thank R. Vipond for assistance with validation experiments. We thank H. Eno and B. Myers for help with proofreading. We are thankful for the generous support of reagents and technical support from Oxford Nanopore. We thank the staff at Oxford Nanopore for technical and logistical support during this project with special thanks to S. Brooking, O. Hartwell, R. Pettett, C. Brown, G. Sanghera and R. Ronan. We thank T. Bedford and R. Neher for developing the Nextstrain website.

Author Contributions N.J.L., J.Q., M.K.O'S., D.W., S.G., M.W.C. conceived the study. N.J.L., J.Q., M.K.O'S., S.A.W., J.T., P.R., D.T. designed the lab in a suitcase and laboratory protocol and initial validation. J.Q., S.D., L.C., J.A.B., R.K., L.E.K., and A.Ma. performed MinION sequencing. N.J.L., J.Q. and J.T.S. performed bioinformatics analysis and wrote software. J.T.S. added variant calling support to the nanopolish software. N.J.L., J.Q., S.D., E.S., P.F., L.C., A.Mi., N.M. and I.R. analysed the data. G.D., A.R., N.J.L., J.Q. and G.P. performed phylogenetic analysis. J.A.H., D.A.M., G.P., K.L., B.A. assisted further validation experiments. M.W.C., M.Ga., S.G., A.D.C., K.S., E.F. and R.W. coordinated activities for the European Mobile Laboratories. N.J.L., J.Q., S.D., M.W.C., S.G., M.K.O'S., A.R., E.S., P.F., I.R., A.Mi., and L.C. wrote the manuscript. All other authors were involved either in sample collection, and/or logistical support and strategic oversight for the work.

Author Information MinION and Illumina raw sequence files have been deposited into the European Nucleotide Archive under project code PRJEB10571. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.J.L. (n.j.loman@bham.ac.uk).

© 2016 Macmillan Publishers Limited. All rights reserved

METHODS

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

Transportation. All equipment was loaded into a Pelican 1610 case (Pelican, Torrance, USA), cold chain reagents were packed into two polystyrene boxes with either ice or cool packs. These were sealed and placed in a holdall with the plastic consumables. Both pieces of luggage were flown by air as normal checked baggage.

RNA extraction. RNA was extracted from 50 μl whole blood, 140 μl serum, 140 μl of resuspended swab or 140 μl urine using the QIAamp Viral RNA Mini Kit (Qiagen, Manchester, UK), following the manufacturer's instructions. Samples were inactivated by adding 560 μl of Buffer AVL (Qiagen) and 560 μl of 100% ethanol while still in a glove box, this method has been shown to inactivate EBOV in blood samples[20]. Following inactivation, samples were handled on the bench employing standard laboratory safety precautions.

RT–PCR. Individual 25 μl RT–PCR reactions were performed using the SuperScript III One-Step RT–PCR System with Platinum Taq DNA Polymerase (Life Technologies Ltd, Paisley, UK). Each reaction was made up by adding 12.5 μl 2 × reaction mix, 1 μl enzyme mix, 1 μl primers (10 μM), 0.5 μl RNA extract and nuclease-free water. Thermocycling was performed on an Eppendorf Master Cycler Personal instrument with the following program: 60 °C for 30 min, 94 °C for 2 min followed by 45 cycles of 94 °C for 15 s, 55 °C for 30 s, 68 °C for 2 min and a final extension of 68 °C for 5 min.

MinION library preparation. Each reaction was quantified on a Qubit 3.0 fluorimeter using the dsDNA HS assay (Life Technologies). Equimolar amounts of each amplicon product to a total DNA mass of 1 μg was pooled into a single tube and cleaned-up using an equal volume of MAGBIO HighPrep PCR beads (AutoQ Biosciences, Reading, UK). Pooled amplicons were diluted to 85 μl, and end-repaired in a total volume of 100 μl, using the NEBNext End Repair Module (New England Biolabs, Hitchin, UK) before being cleaned up using an equal volume of HighPrep PCR beads and eluting in 25 μl nuclease-free water. 3′ dA-tailing was performed using the NEBNext dA-Tailing Module (New England Biolabs) in a volume of 30 μl, before being cleaned up using an equal volume of HighPrep PCR beads and eluting in 30 μl nuclease-free water. 10 μl of 'Adaptor mix' and 10 μl 'HP adaptor' supplied in the SQK-MAP005 library preparation kit (Oxford Nanopore Technologies, Oxford, UK) were added to the dA-tailed amplicons along with 50 μl, Blunt/TA Ligase Master Mix (New England Biolabs) in a Protein LoBind tube (Eppendorf UK) and incubated for 10 min. The resulting sequencing library was purified using Dynabeads His-Tag Isolation and Pulldown beads (Life Technologies, Stevenage, UK) according the SQK-MAP005 protocol supplied by Oxford Nanopore Technologies as part of the MinION Access Program. The final library was quantified using the Qubit to confirm the process had been successful. 6 μl, of library was diluted using 75 μl '2x Running Buffer', 66 μl Nuclease-free water (Promega UK, Chilworth, UK) and 3 μl, 'Fuel Mix'.

MinION sequencing. A new flowcell was unpackaged and fitted onto the MinION device. The flowcell was primed with a blank sample created as described above, and left to incubate for 10 min. The priming process was repeated a second time before the sample was loaded. Running MinKNOW version 0.49.2.9 and starting the protocol 'MAP_48Hr_Sequencing_Run.py' initiated the sequencing run. An offline-capable version of MinKNOW, with internet 'ping' disabled and online updates disabled was made available to us by Oxford Nanopore Technologies specifically for the project (available on request from Oxford Nanopore Technologies).

Data transfer. With no method of offline analysis available during the outbreak period, there was a dependency on local internet connectivity to facilitate the upload of the raw FAST5 files produced by MinKNOW. A variety of methods were used depending on location and circumstances with the vast majority of the data being uploaded from the European Mobile Laboratories staff accommodation in Coyah, Guinea, via a mobile internet 3G hotspot (TP-LINK M5350 3G hotspot on the MTN mobile network). At times due to unknown factors the upload speed was limited to 2G and took significantly longer. Using Cygwin version 2.0.0 and the Linux tar command a compressed archive containing the first 5,000 to 10,000 .fast5 read files generated by each run was created. This was uploaded to a Google Drive shared directory. Eventually in Coyah we were provided access to a broadband connection (MTN network, 5 Mb s^{-1}, established by the World Food Program), which proved to be more reliable than mobile internet.

Data handling. Data was downloaded onto a Linux server on the MRC Cloud Infrastructure for Microbial Bioinformatics located in Birmingham, UK. Files were unpacked and basecalled using the Metrichor command-line interface and the workflow 2D Basecalling for MAP-005 (versions 1.14, 1.24 and 1.34). This software was provided by Oxford Nanopore Technologies (available on request) for the project in order to permit basecalling to be carried out through the Linux command line as part of a pipeline. The MinION generates one direction (1D) and two direction reads (2D). 2D reads are higher quality and were used for analysis. 2D reads that were in the pass filter folder and 2D reads designated as high-quality (due to having more complement events than template events) in the fail folder as determined by poretools were extracted into FASTA (for nanopolish) and FASTQ format (for marginAlign) with poretools version 0.5.1[21].

Bioinformatics analysis. We use a reference mapping approach to detect single nucleotide variants through alignment to a reference strain from early in the outbreak (GenBank accession number EM_079517)[11]. Due to the nature of the sequencing data, which is dominated by insertion and deletion errors, we do not attempt to call insertion or deletions[14]. Variants were detected using the variants module of the nanopolish software package. Initial nucleotide base alignment was carried out with MarginAlign[12]. Nanopolish then uses the event-level ('squiggle') data generated by the MinION to evaluate candidate variants found in the aligned reads as described in the following section. Variants with a log likelihood ratio of >200 and coverage depth of >50× (25× 2D coverage) are accepted and a consensus sequence is generated for each sample. Regions of uncertainty (for example in difficult to sequence homopolymeric regions or primer binding sites), or with low coverage (<50×, or 25× 2D coverage) are masked with an N character. Assuming sufficient genomic coverage is present over a specific amplified variant this approach gives a high true positive variant calling rate. However, failure of individual amplicons to amplify, or unbalanced coverage of regions may reduce this figure. This is assessed, on each individual sample, by artificially mutating the reference genome with 30 randomly chosen mutations. Mutated positions in the references should be detected as variants, using the simplifying assumption that these variants are unlikely to be present in the sample. Any positions not covered by the tiling amplicon scheme (that is, the extreme 5′ and 3′ ends) are not considered in the true positive rate calculation. Each sample is therefore assigned a quality indicator. Those with a true positive rate (TPR, that is, sensitivity) of ≥75% are included in phylogenetic inferences. Samples with TPR <75% were not used for the phylogenetic analysis presented here.

Signal-based SNP calling. SNPs were called using the "variants" module from the nanopolish package (manuscript in preparation, https://github.com/jts/nanopolish, branch snp_calling_alternative_models, commit ID 25ea7bac3ab9e1d266079ac105ab2005cfa39a14).

The nanopolish variants program first finds candidate SNPs by finding mismatches between the aligned nanopore reads and the reference genome. These candidate SNPs are clustered into sets of nearby SNPs, an exhaustive set of candidate haplotypes are derived from the possible combinations of SNPs and the haplotype that maximizes the probability of the event-level data called as the sequence for region. We describe each step in detail below.

Candidate SNP generation. We iterate over the entire reference genome and examine positions covered by at least 20 nanopore reads. At these well-covered positions we considered any non-reference base that was seen in at least 20% of the nanopore reads to be a candidate SNP. These candidates were passed to the next stage of the pipeline.

Candidate haplotype generation. As the MinION sequencer does not measure single bases, but rather current signals dependent on a short sequence of nucleotides that are in the pore, we could not assess each SNP individually. Instead, we partitioned the set of candidate SNPs into groups whose signals may interact and overlap. We determined that SNPs separated by at least 10 bp could be treated independently; therefore we partitioned the candidate SNP set into subsets of SNPs that are within 10 bp of each. For each subset of candidate SNPs we exhaustively generated all possible haplotype sequences by including/excluding the individual SNPs in the subset. As the number of possible combinations of n SNPs is 2^n, we had to discard subsets that contained more than 10 candidate SNPs or spanned a reference region greater than 100 bp. For each derived haplotype sequence S, we calculate the likelihood of S using a modified version of the hidden Markov model (HMM) we previously described[16].

Haplotype likelihoods. The nanopolish HMM calculates the probability of observing a sequence of events emitted by the nanopore, which we denote as D, given an arbitrary sequence S. The structure of the HMM is as previously described but now allows events to be "soft-clipped" to better handle uncertainty about where the event-to-sequence alignment starts and ends. In addition, we incorporated a new model from Oxford Nanopore that models the event signals to be dependent on six-base-pair subsequences rather than five-base-pair subsequences. To use this model on SQK-MAP-005 data we calculated a global shift parameter (shift_offset) that rescales SQK-MAP-005 data to the 6 bp emission functions. We otherwise did not train the emission functions, per-read scaling parameters or transition probabilities of our hidden Markov model.

© 2016 Macmillan Publishers Limited. All rights reserved

Variant calls. For each subset of candidate SNPs, the haplotype with the largest likelihood is called as the sequence for the region. The SNPs contained on the called haplotype (if any) are output in VCF format. The log likelihood ratio between the called haplotype and the reference haplotype (containing no SNPs) was output as the score for each variant to facilitate downstream filtering. Metadata such as the total depth of the region and the number of reads that support the called haplotype over the reference sequence is also output.

Validation experiments. *Dstl amplicons.* Archived Zaire Ebolavirus was amplified using 38 primer pairs, giving approximately 500 base pair amplicons, according to the study protocol. As this work was before in-field sequencing, different versions of the MinKNOW software and Metrichor basecaller were used. Amplicons were sequenced by both MinION and Illumina. An Illumina library was constructed from the same amplicon pool and tagmented using the Nextera XT library preparation kit. The library was sequenced on the Illumina MiSeq. Because of the huge excess of coverage generated, this data set was subsampled to 400,000 paired reads before aligning to the EM_079517 reference sequence using BWA-MEM[22]. After sorting and converting the resulting alignment to BAM using samtools, variants were determined using FreeBayes[23]. A consensus sequence was generated using the vcf2fasta component of vcflib (https://github.com/ekg/vcflib). The MinION data was analysed as per the study methods, except for a modification to nanopolish to allow it to consider up to 15 variants per segment in order to account for the increased divergence between the genome and the reference. The MinION and Illumina consensus sequencs were aligned using the nucmer component of MUMMER and variants determined using the show-snps module[24]. Scripts and documentation for this analysis are in the Github notebook Dstl validation.ipynb.

180 genome analysis. Six samples of leftover RNA from a previously performed sequencing study[3] were processed at Public Health England Porton Down, as per the methods described in the manuscript. One sample did not yield any sequenceable products, so five genomes (EM_076534, EM_076533, EM_076383, EM_078416, EM_076769) were sequenced on MinION at PHE Porton Down. The 11 reaction scheme was used except for sample EM_076769 when the 19 reaction scheme was used. These sequences were compared with Illumina consensus sequences from the previously published data set in Carroll *et al.*[3]. Variants were identified between the reference genome (EM_079517) and each of the successfully samples using the show-snps component of MUMMER[24]. Variants detected by our pipeline were compared against expected variants, before and after quality filtering, using custom Python scripts deposited in the Github repository and documented in the IPython Notebook. A phylogeny was inferred using RaXML[25] including the consensus sequences from the validation set along with all of the consensus sequences from Carroll *et al.*[3] MinION sequence accuracy rates for two-direction (2D) reads were determined using A. Quinlan's count-errors.py script (http://github.com/arq5x/nanopore-scripts) as described in Quick *et al.*[11]. Scripts and documentation for this analysis are in the Github notebook: Examine validation runs.ipynb.

Analysis of SNP calling sensitivity. Reads were subsampled at collection time intervals using the poretools times command[21], simulating the order reads are obtained by real-time sequencing on the nanopore, to demonstrate the effect of coverage on SNP calling sensitivity and log likelihood ratio.

Analysis of samples from the same patient. Samples were analysed as part of the real-time surveillance work. The consensus sequences from four pairs of samples each from four individuals were generated. Each pair was compared individually using the show-snps module of MUMmer to investigate differences.

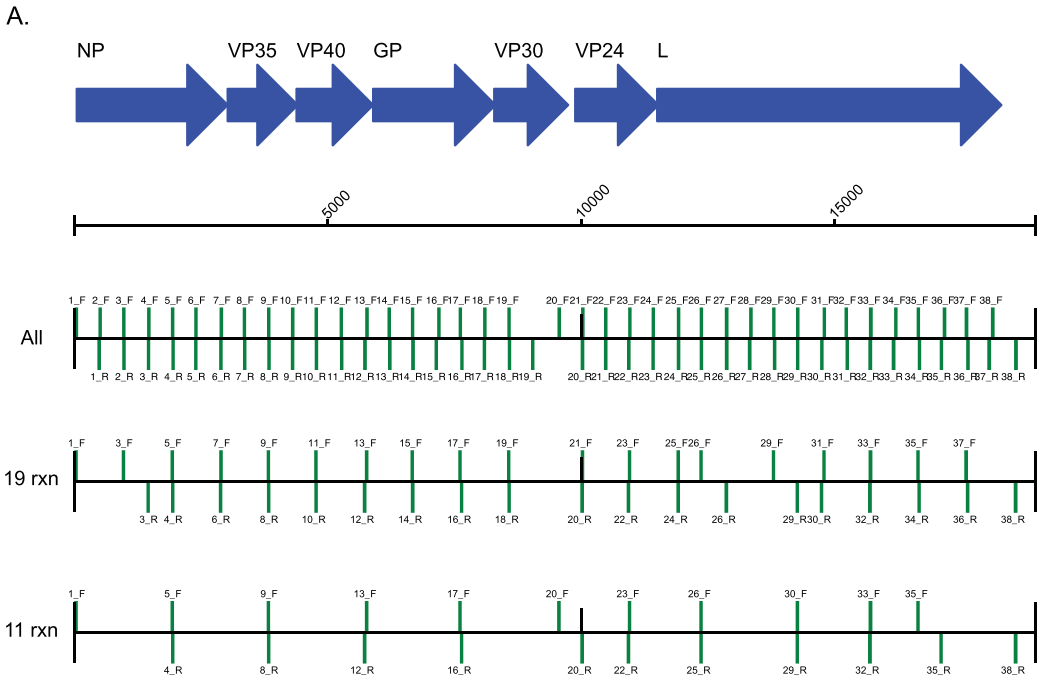
Detection of putative transmission events from Sierra Leone. We downloaded the 74 genome sequences made available on http://virological.org (http://virological.org/t/direct-deep-sequencing-in-sierra-leone-yields-73-new-ebov-genomes-from-february-may-2015/134) and aligned them against sequences from our analysis using MUSCLE[26]. We then generated a phylogenetic tree using FastTree 2 with the GTR model[27]. Any sequences that fell into the GN1 or SL3 lineages were included in future analysis.

Phylogenetic inferences. Consensus sequences from real-time sequencing were aligned with previously published genome sequences from Guinea, Sierra Leone and Liberia[7]. To address the over-representation of Sierra Leone sequences in this set we randomly down-sampled available sequences, resulting in a total of 313 sequences from Sierra Leone. Maximum likelihood trees are produced using RAxML 8.2.3 using the GTRGAMMA model and 200 bootstrap replicates[25]. Time-scaled trees were produced with BEAST v1.8.2[28] using a HKY+gamma substitution model[29,30] partitioned by first, second and third codon positions and intergenic regions, a Skygrid tree prior[31] and an uncorrelated lognormal clock[32], and an uninformative prior on the mean of the molecular clock rate (XML file in the accompanying Github repository). The maximum clade credibility tree was recovered using TreeAnnotator. Phylogenetic trees were annotated using the ete3 Python package.

Code availability. Reproducible workflows for the analysis presented here and consensus sequences can be found at http://github.com/nickloman/ebov and are freely available under the MIT license. The complete set of bioinformatics scripts are available in a Github repository with associated IPython Notebooks to regenerate the figures and tables presented in this manuscript can be found at http://github.com/nickloman/ebov.

20. Smither, S. J. *et al.* Buffer AVL alone does not inactivate Ebola virus in a representative clinical sample type. *J. Clin. Microbiol.* **53,** 3148–3154 (2015).
21. Loman, N. J. & Quinlan, A. R. Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics* **30,** 3399–3401 (2014).
22. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at http://arxiv.org/abs/1303.3997 (2013).
23. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. Preprint at http://arxiv.org/abs/1207.3907 (2012).
24. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5,** R12 (2004).
25. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–1313 (2014).
26. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
27. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* **5,** e9490 (2010).
28. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7,** 214 (2007).
29. Hasegawa, M., Kishino, H. & Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22,** 160–174 (1985).
30. Yang, Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39,** 306–314 (1994).
31. Gill, M. S. *et al.* Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30,** 713–724 (2013).
32. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4,** e88 (2006).
33. Cock, P. J. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25,** 1422–1423 (2009).

© 2016 Macmillan Publishers Limited. All rights reserved



B. 19 reactions

Forward	Reverse	Length
1_F	3_R	1426
3_F	4_R	973
5_F	6_R	952
7_F	8_R	941
9_F	10_R	940
11_F	12_R	958
13_F	14_R	906
15_F	16_R	974
17_F	18_R	969
19_F	20_R	1445
21_F	22_R	906
23_F	24_R	958
25_F	26_R	947
26_F	29_R	1898
29_F	30_R	946
31_F	32_R	901
33_F	34_R	963
35_F	36_R	977
37_F	38_R	975

C. 11 reactions

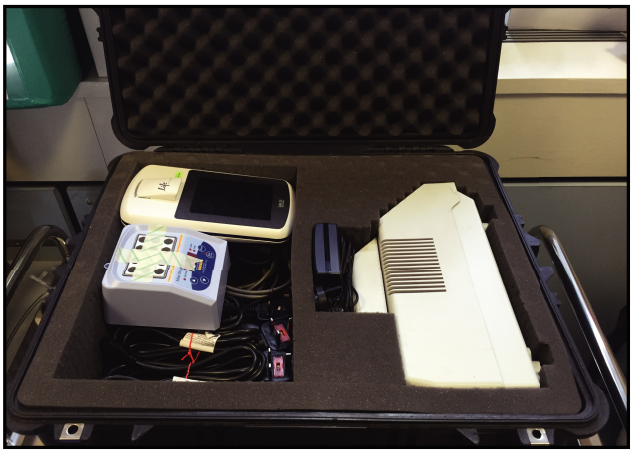
Forward	Reverse	Length
1_F	4_R	1911
5_F	8_R	1901
9_F	12_R	1895
13_F	16_R	1874
17_F	20_R	2406
20_F	22_R	1371
23_F	25_R	1410
26_F	29_R	1898
30_F	32_R	1427
33_F	35_R	1396
35_F	38_R	1921

Extended Data Figure 1 | Primer schemes employed during the study. We designed PCR primers to generate amplicons that would span the EBOV genome. **a**, We initially designed 38 primer pairs which were used in the initial validation study and which cover >97% of the EBOV genome. During in-field sequencing we used a 19-reaction scheme or 11-reaction scheme, which generated longer products. The predicted amplicon products are shown with forward primers and reverse primers indicated by green bars on the forward and reverse strand, respectively, scaled according to the EBOV virus coordinates. **b**, **c**, The amplicon product sizes expected are shown for the 19-reaction scheme (**b**) and the 11-reaction scheme (**c**). No amplicon covers the extreme 3′ region of the genome. The last primer pair, 38_R, ends at position 18578, 381 bases away from the end of the virus genome. The primer diagram was created with Biopython[33].

© 2016 Macmillan Publishers Limited. All rights reserved

A. Equipment

Item	Number	Model
Thermocycler	1-3	MasterCycler Personal (Eppendorf)
Fluorometer	1	Qubit 3.0 (Life Technologies)
Laptop	2-3	NT310-H (Stone)
MinION	2-3	-
Pipettes	6	P2, 10, 20, 100, 200, 1000 (Gilson)
Microfuge	1-2	
Dry bath	1	Mini Dry Bath Incubator (Starlab)
Magnetic rack	1	MagnaRack (Life Technologies)
Power strip	1	Dependent on country



B. Consumables

Item	Supplier
DNA LoBind Tubes (2 ml)	Eppendorf
Protein LoBind Tubes (2 ml)	Eppendorf
Qubit Assay Tubes	Life Technologies
PCR Tubes with Flat Caps (0.2 ml)	Starlab
Pipette Tips (10 µl, 20 µl, 100 µl, 200 µl, 1000 µl)	Sarstedt
Nitrile Gloves	Kimberly Clark Professional

C. Reagents

Reagent	Shipping Condition	Supplier
Nuclease-Free Water	Ambient	Qiagen
Ethanol 100%	Ambient	-
HighPrep PCR	Chilled	MAGBIO
Dynabeads His-Tag Isolation and Pulldown	Chilled	Life Technologies
Oligos	Chilled	Sigma
Qubit dsDNA HS Assay Kit	Chilled	Life Technologies
MinION Flowcells	Chilled	Oxford Nanopore Technologies
NEBNext End-Repair Module	Frozen	New England Biolabs
NEBNext dA-Tailing Module	Frozen	New England Biolabs
Blunt/TA Ligase Master Mix	Frozen	New England Biolabs
SuperScript III One-Step RT-PCR System with Platinum Taq DNA Polymerase	Frozen	Life Technologies
SQK-MAP005	Frozen	Oxford Nanopore Technologies

Extended Data Figure 2 | List of equipment and consumables to establish the genome surveillance system. a–c, We show the list of equipment (**a**), disposable consumables (**b**) and reagents (**c**) to establish in-field genomic surveillance. Sufficient reagents were shipped for 20 samples. MinION sequencing requires a mix of chilled and frozen reagents. Recommended shipping conditions are specified. The picture underneath depicts MinION flowcells ready for shipping with insulating material (left) and frozen reagents (right).

© 2016 Macmillan Publishers Limited. All rights reserved

Base calling	Convert nanopore squiggles to nucleotide sequences	`metrichor-cli`
Convert to FASTA/FASTQ	Extract basecalled information from nanopore FAST5 files	`poretools fasta --type 2D pass/` `poretools fasta --type 2D --high-quality fail/`
Align to reference	Align sequences to reference	`marginAlign --inputModel input.hmm EM_079517.fasta reads.fastq out.sam`
Local HMM realignment	Iteratively improve alignment based on nanopore insertion/deletion/substitution rates	
Alignment trimming	Remove alignments outside of primer regions in case of adaptor contamination	`align_trim.py`
Event alignment	Map individual event k-mers to reference genome guided by base alignment	`nanopolish eventalign`
Variant calling	Extract candidate mutations from aligned reads, cluster them and evaluate them using a 6-mer HMM	`nanopolish variants`
Consensus building	Mask positions in the genome with either i) <50x 1-D coverage ii) low-quality variants detected iii) in primer binding site	`margin_cons.py`

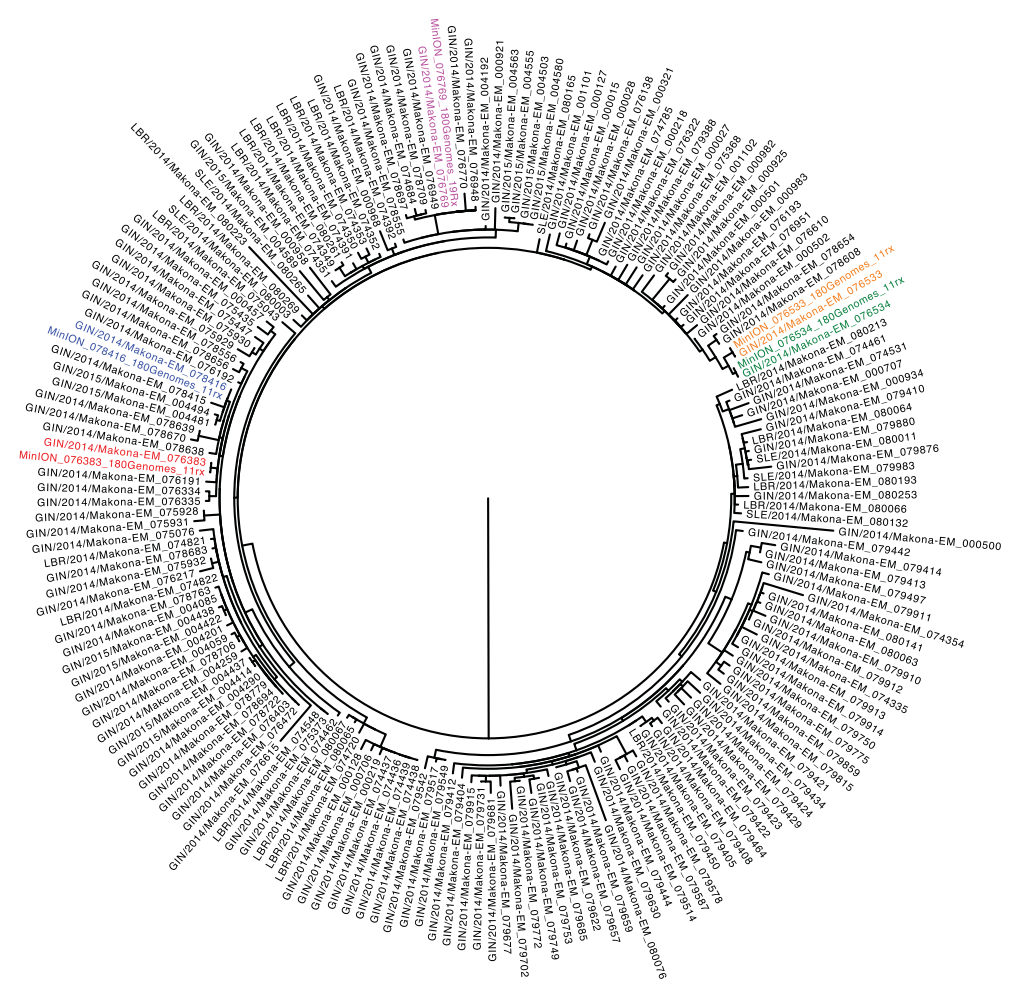
Extended Data Figure 3 | Bioinformatics workflow. This figure summarizes the steps performed during bioinformatics analysis (ordered from top to bottom), in order to generate consensus sequences. The right column shows the example software command executed at each step.

© 2016 Macmillan Publishers Limited. All rights reserved

A

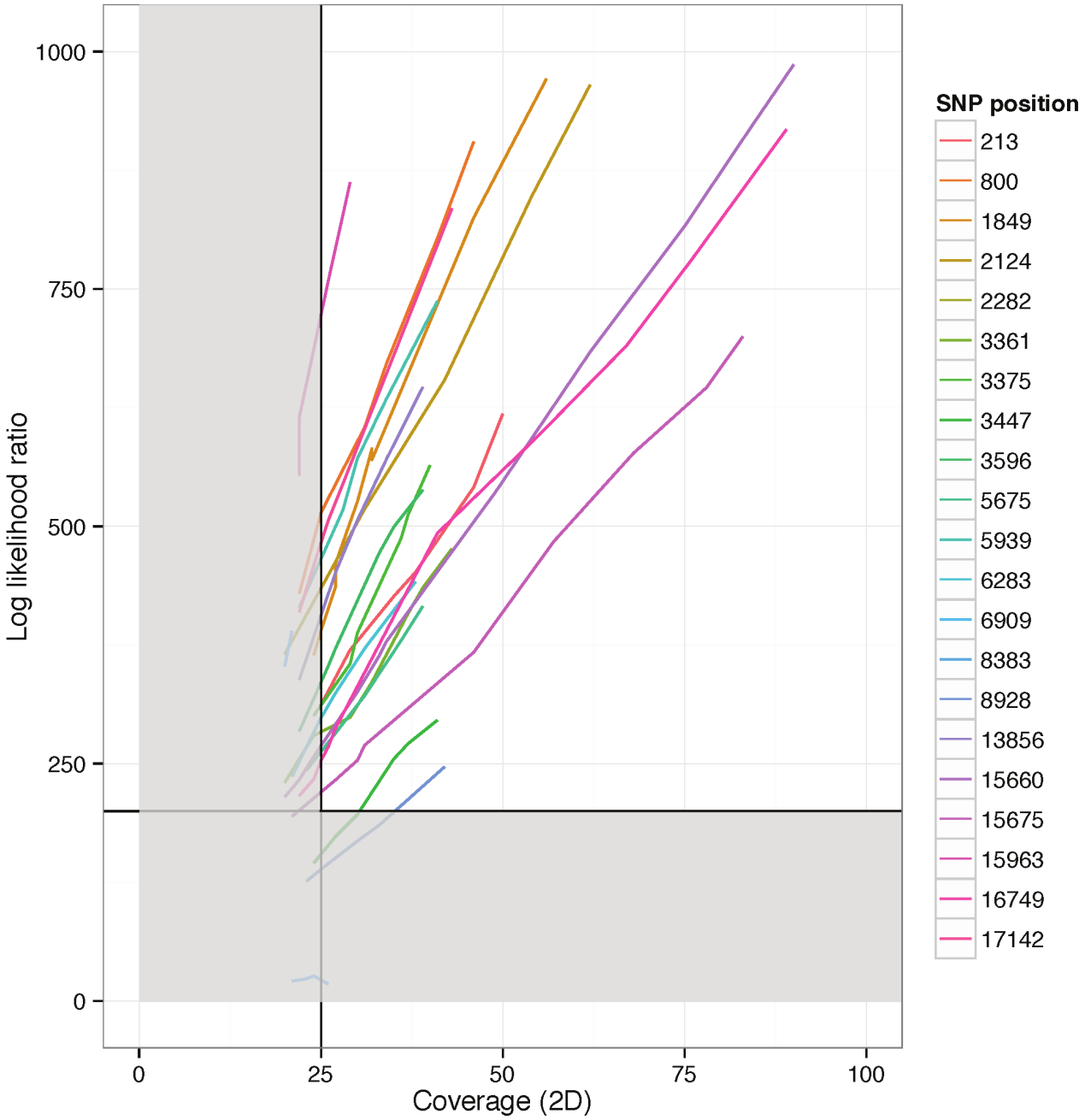
Sample	Reaction scheme	Total mutations	True positives	False positives	False negatives		
					Masked		Wrong
					Outside amplicon scheme	Primer binding site	Erroneous reference call
Pre-quality filtering (all variants)							
076533	11 reactions	20	19	1	1	-	-
076383	11 reactions	18	16	1	1	1	-
078416	11 reactions	18	17	1	-	1	-
076769	19 reactions	19	19	0	-	-	-
Post-quality filtering (high quality variants)							
076533	11 reactions	20	19	0	1	-	-
076383	11 reactions	18	16	0	1	1	-
078416	11 reactions	18	17	0	-	1	-
076769	19 reactions	19	19	0	-	-	-

B



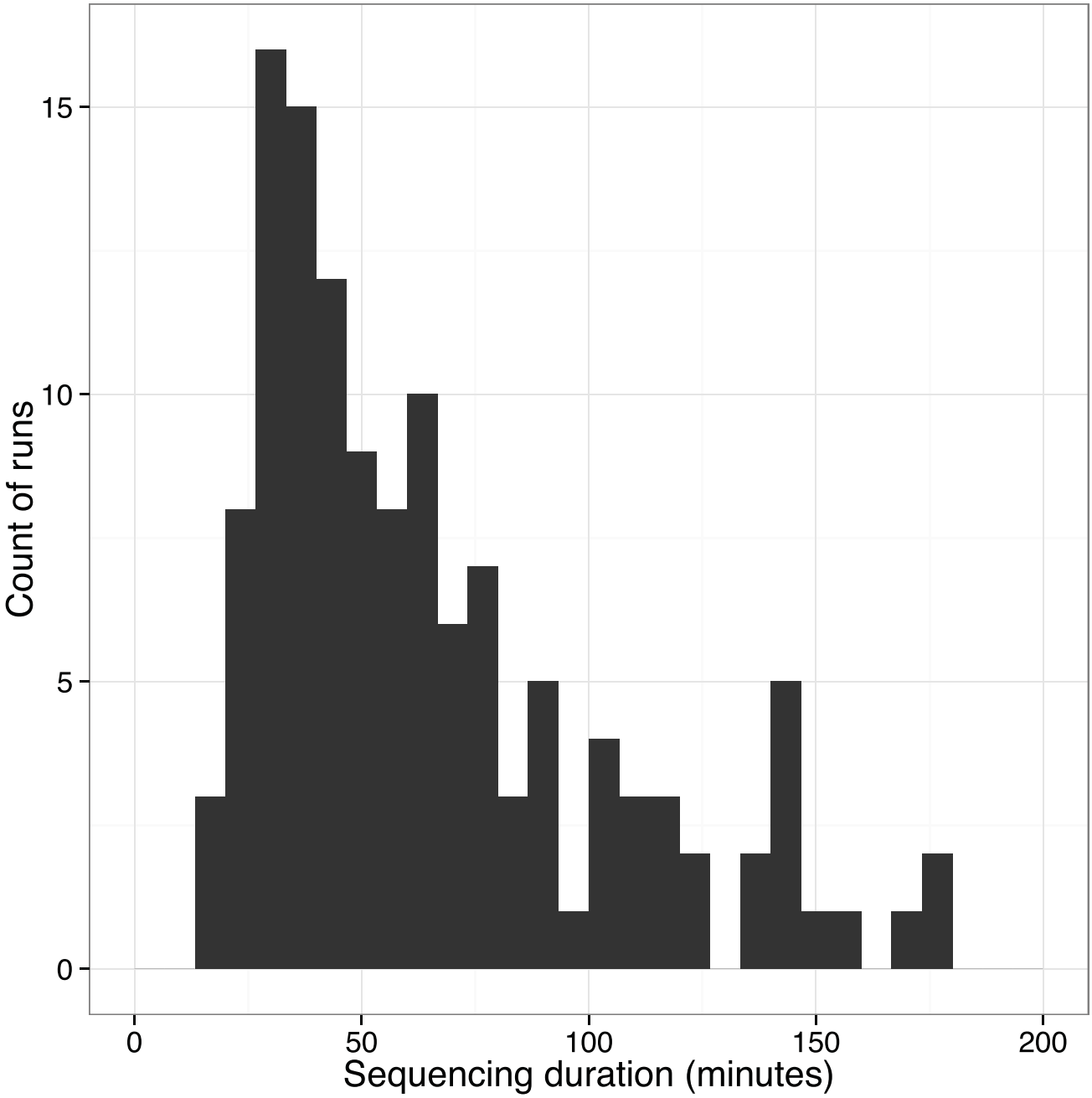
Extended Data Figure 4 | Results of MinION validation. a, The results of comparing four MinION sequences with Illumina sequences generated as part of a previous study[3] are shown. Each row in the table demonstrates the number of true positives, false positives and false negatives for a sample. False negatives may result in masked sequences, owing to being outside of regions covered by the amplicon scheme, having low coverage or falling within a primer binding site. Results before and after quality filtering (log likelihood ratio of >200) are shown. After quality filtering, no false positive calls were detected. All detected false negatives were masked with Ns in the final consensus sequence. No positions were called incorrectly. **b**, The four consensus sequences, plus an additional sample that had missing coverage in one amplicon are shown as part of a phylogenetic reconstruction with genomes from Carroll *et al.*[3]. Sample labels in red, blue, pink, yellow and blue represent pairs of sequences generated on MinION and llumina. These fall into identical clusters.

© 2016 Macmillan Publishers Limited. All rights reserved



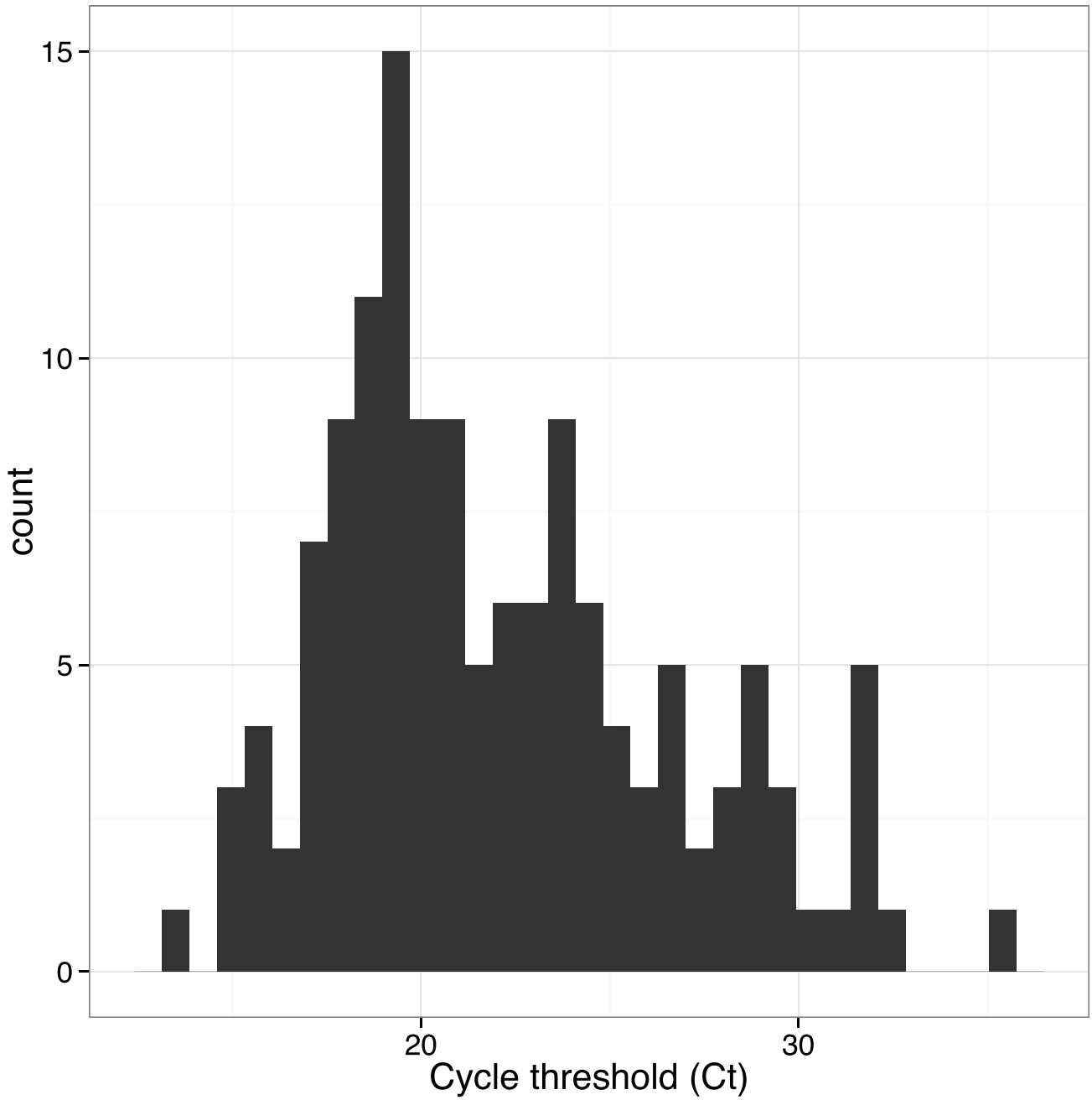
Extended Data Figure 5 | Relationship between coverqage and log-likelihood ratio for sample 076769. Line-plot showing the relationship between sequence depth of coverage (*x* axis) and the log likelihood ratio for detected SNPs derived by subsampling reads from a single sequencing run to simulate the effect of low coverage. The horizontal and vertical line indicates the cut-offs (quality and coverage respectively) for consensus calling. Therefore, all variants are detected below 25× coverage, and the vast majority meet the threshold quality at 25× coverage or slightly above. Any combination of log likelihood ratio or coverage that placed variants in the grey box would be represented as a masked position in the final consensus sequence.

© 2016 Macmillan Publishers Limited. All rights reserved



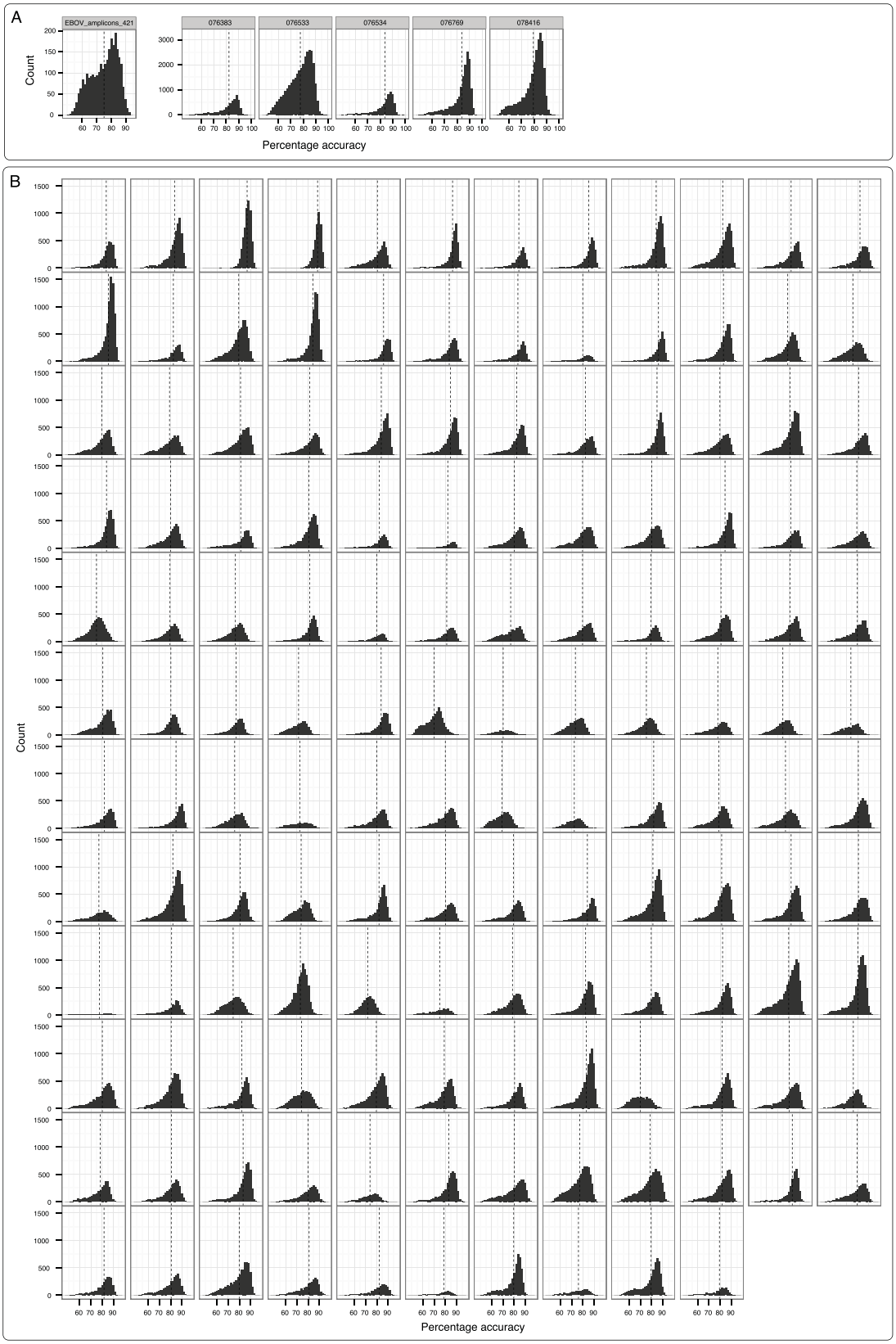
Extended Data Figure 6 | Duration of MinION sequencing runs. For each sequence run the sequencing duration, measured as the difference between timestamp of the first read seen and the last read transferred for analysis. 127 runs are shown, with 15 outliers with duration greater than 200 min excluded.

© 2016 Macmillan Publishers Limited. All rights reserved



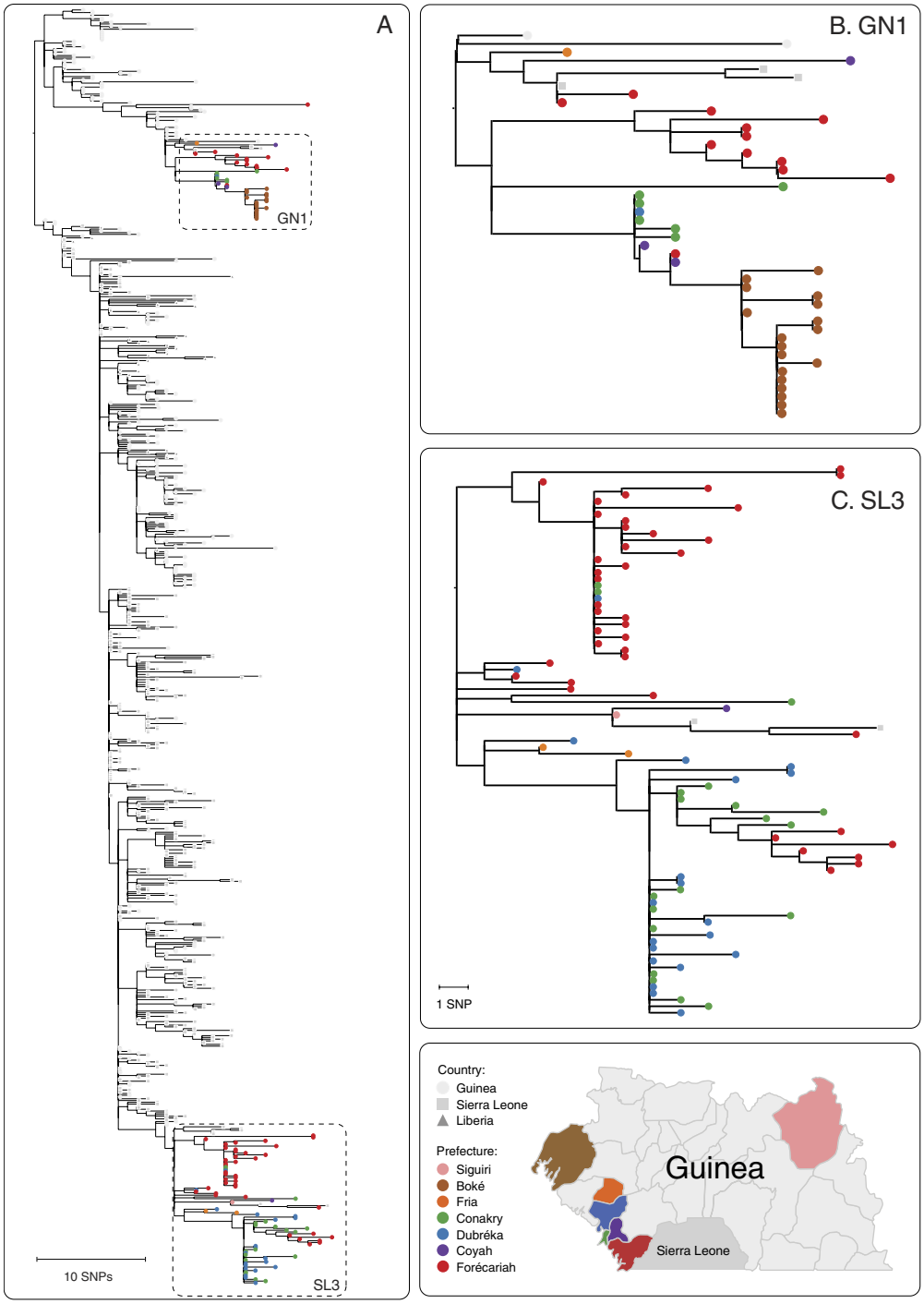
Extended Data Figure 7 | Histogram of C_t values for study samples. C_t values for samples in the study (where information was available) ranged between 13.8 and 35.7, with a mean of 22.

© 2016 Macmillan Publishers Limited. All rights reserved



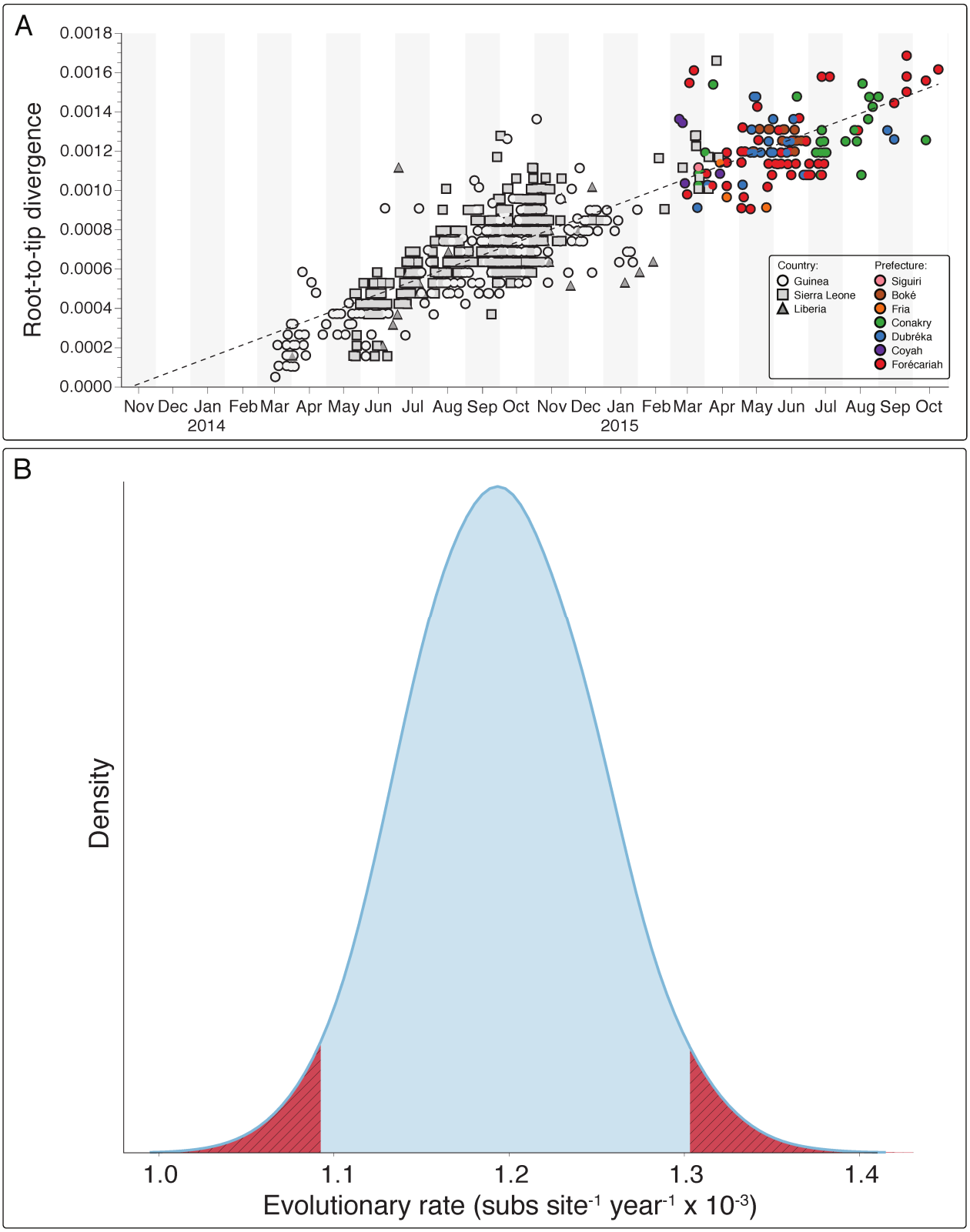
Extended Data Figure 8 | Sequence accuracy for samples. a, **b**, Accuracy measurements for the entire set of two-direction reads were made for the validation samples, sequenced in the United Kingdom (**a**) and each of the 142 samples from real-time genomic surveillance (**b**). Accuracy is defined according to the definition from Quick *et al.*[11]. Vertical dashed lines indicate the mean accuracy for the sample.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 9 | Maximum likelihood phylogenetic inference of 125 Ebola virus samples from this study with 603 previously published sequences. Coloured nodes are from this study. Node shape reflects country of origin. **a**–**c**, the entire data set is shown (**a**), with zoomed regions focusing on lineages GN1 (**b**) and SL3 (**c**) identified during real-time sequencing. Map figure adapted from SimpleMaps website (http://simplemaps.com/resources/svg-gn).

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 10 | Root-to-tip divergence plot and mean evolutionary rate estimate. a, Root-to-tip divergence plot for the 728 Ebola samples generated through maximum likelihood analysis. Samples from real-time genomic surveillance are coloured as per Fig. 3 and Extended Data Fig. 9. **b**, Mean evolutionary rate estimate (in substitutions per site per year) across the EBOV phylogeny recovered using BEAST under a relaxed lognormal molecular clock. Blue area corresponds to the 95% highest posterior density (HPD) (mean of the distribution is 1.19×10^{-3}, 95% HPDs: $1.09–1.29 \times 10^{-3}$ substitutions per site per year). Hatched regions in red are outside the 95% HPD intervals.

© 2016 Macmillan Publishers Limited. All rights reserved