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. 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 $0.87 \times 10^{-3}$ and $1.42 \times 10^{-3}$ 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. 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. 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. 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 sequencing instruments, such as those using optical
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 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. 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
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).
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 leftover 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 studies2–7. 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 Leone3, 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 20144. Through integration of our data set with those generated by 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 work18. By contrast, the MinION sequencing 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 persist15.

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://www.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 individuals19 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/liberas-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 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://bioxiv.org/content/early/2015/03/13/15281 (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).
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. Mata, 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 nanoplanish 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. 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).
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. 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 nuclelease-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.

MiniION 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 MAGE BIO 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 nuclelease-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 nuclelease-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 Pulpdown beads (Life Technologies, Stevenage, UK) according the SQK-MAP005 protocol supplied by Oxford Nanopore Technologies as part of the MiniION 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 Nuclelease-free water (Promega UK, Chilworth, UK) and 3 μl ‘Fuel Mix.’

MiniION sequencing. A new flowcell was unpackaged and fitted onto the MiniION 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 uploading of raw FAST5 read 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 MiniION 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 nanoplan) and FASTQ format (for marginAlign) with poretools version 0.5.12. 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)17. Due to the nature of the sequencing data, which is dominated by insertion and deletion errors, we do not attempt to call insertion or deletions14. Variants were detected using the variants module of the nanoplan software package. Initial nucleotide base alignment was carried out with MarginAlign12. Nanoplan then uses the event-level (‘squiggle’) data generated by the MiniION 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 a 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 filling 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 nanoplan package (manuscript in preparation, https://github.com/jfs/napopolish, branch snp_calling_alternative_models, commit ID 25a7a/bac3ab9e1d266079ac 105a2b6050fca39a14). The nanoplan variant programs first find 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 MiniION 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 overlapped. 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 2ⁿ, we have found that 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 described16.

Haplotype likelihoods. The nanoplan 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 positions of 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 6bp emission functions. We otherwise did not train the emission functions, per-reading parameters or transition probabilities of our hidden Markov model.
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-MEM22. After sorting and converting the resulting alignment to BAM using samtools, variants were determined using FreeBayes23. A consensus sequence was generated using the vcf2fasta component of vcflib (https://github.com/nickloman/vcflib). The MinION data was analysed as per the study methods, except for a modification to nanoplot 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 sequences were aligned using the nucleotide component of MUMMER and variants determined using the show-snps module24. 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 study27 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_076533, 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 MUMMER24. 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 RaXML25 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/arqxs/nanopore-scripts) as described in Quick et al.11. Scripts and documentation for this analysis are in the Github notebooks: Examine validation runs.ipynb.

Analysis of SNP calling sensitivity. Reads were subsampled at collection time intervals using the poretools times command13, 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 MUSCLE26. We then generated a phylogenetic tree using FastTree2 with the GTR model32. 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 genotypes with sequences from Guinea, Sierra Leone and Liberia3. 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 GTR+Gamma model and 200 bootstrap replicates25. Time-scaled trees were produced with BEAST v1.8.226 using a HKY+gamma substitution model29,30 partitioned by first, second and third codon positions and intergenic regions, a Skygrid tree prior31 and an uncorrelated lognormal clock32, 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/ebow.

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. Liu, J., Lan, 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. 30, 306–314 (1990).
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
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\textsuperscript{33}. 

| Reaction Scheme | Primer Pairs | Length |
|-----------------|--------------|--------|
| 19 rxn          | 228          | 4640   |
| 11 rxn          | 57           | 2000   |

\textsuperscript{33} Biopython is a free and open-source software library for computational molecular biology and biinformatics.
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).
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.
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 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. Sample labels in red, blue, pink, yellow and blue represent pairs of sequences generated on MinION and Illumina. These fall into identical clusters.
Extended Data Figure 5 | Relationship between coverage 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.
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.
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.
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. Vertical dashed lines indicate the mean accuracy for the sample.
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 $1.19 \times 10^{-3}$, 95% HPDs: $1.09-1.29 \times 10^{-3}$ substitutions per site per year). Hatched regions in red are outside the 95% HPD intervals.