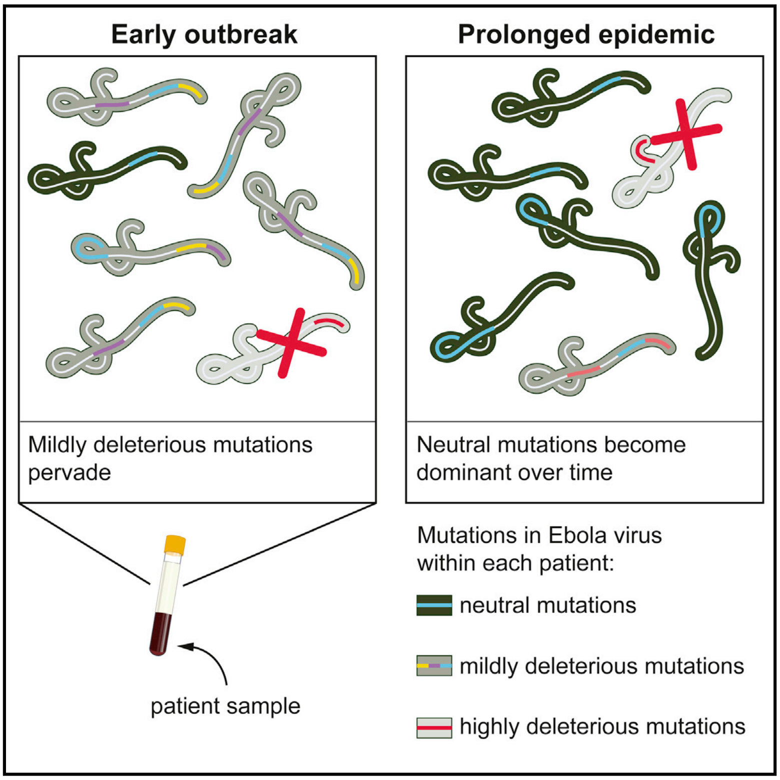
Cell

Article

Ebola Virus Epidemiology, Transmission, and Evolution during Seven Months in Sierra Leone

Graphical Abstract


Early outbreak
Prolonged epidemic
Mildly deleterious mutations pervade
Neutral mutations become dominant over time
patient sample
Mutations in Ebola virus within each patient:
neutral mutations
mildly deleterious mutations
highly deleterious mutations

Highlights

- In Sierra Leone, transmission has primarily been within-country, not between-country
- Infectious doses are large enough for intrahost variants to transmit between hosts
- A prolonged epidemic removes deleterious mutations from the viral population
- There is preliminary evidence for human RNA editing effects on the Ebola genome

Authors

Daniel J. Park, Gytis Dudas, Shirlee Wohl, ..., Andrew Rambaut, Robert F. Garry, Pardis C. Sabeti

Correspondence

dpark@broadinstitute.org (D.J.P.),
a.rambaut@ed.ac.uk (A.R.),
pardis@broadinstitute.org (P.C.S.)

In Brief

Ebola virus genomes from 232 patients sampled over 7 months in Sierra Leone were sequenced. Transmission of intrahost genetic variants suggests a sufficiently high infectious dose during transmission. The human host may have caused direct alterations to the Ebola virus genome.

Park et al., 2015, Cell *161*, 1516–1526
June 18, 2015 ©2015 The Authors
http://dx.doi.org/10.1016/j.cell.2015.06.007

CellPress

Ebola Virus Epidemiology, Transmission, and Evolution during Seven Months in Sierra Leone

Daniel J. Park,[1,21,*] Gytis Dudas,[2,21] Shirlee Wohl,[1,3,21] Augustine Goba,[4,21] Shannon L.M. Whitmer,[5,21] Kristian G. Andersen,[6] Rachel S. Sealfon,[1,7] Jason T. Ladner,[8] Jeffrey R. Kugelman,[8] Christian B. Matranga,[1] Sarah M. Winnicki,[1,3] James Qu,[1] Stephen K. Gire,[1,3] Adrianne Gladden-Young,[1] Simbirie Jalloh,[4] Dolo Nosamiefan,[1] Nathan L. Yozwiak,[1,3] Lina M. Moses,[9] Pan-Pan Jiang,[1,3] Aaron E. Lin,[1,3] Stephen F. Schaffner,[1,3] Brian Bird,[5] Jonathan Towner,[5] Mambu Mamoh,[4] Michael Gbakie,[4] Lansana Kanneh,[4] David Kargbo,[4] James L.B. Massally,[4] Fatima K. Kamara,[4] Edwin Konuwa,[4] Josephine Sellu,[4] Abdul A. Jalloh,[4] Ibrahim Mustapha,[4] Momoh Foday,[4] Mohamed Yillah,[4] Bobbie R. Erickson,[5] Tara Sealy,[5] Dianna Blau,[5] Christopher Paddock,[5] Aaron Brault,[5] Brian Amman,[5] Jane Basile,[5] Scott Bearden,[5] Jessica Belser,[5] Eric Bergeron,[5] Shelley Campbell,[5] Ayan Chakrabarti,[5] Kimberly Dodd,[5] Mike Flint,[5] Aridth Gibbons,[5] Christin Goodman,[5] John Klena,[5] Laura McMullan,[5] Laura Morgan,[5] Brandy Russell,[5] Johanna Salzer,[5] Angela Sanchez,[5] David Wang,[5] Irwin Jungreis,[7] Christopher Tomkins-Tinch,[1] Andrey Kislyuk,[10] Michael F. Lin,[10] Sinead Chapman,[1] Bronwyn MacInnis,[1] Ashley Matthews,[1,3] James Bochicchio,[1] Lisa E. Hensley,[11] Jens H. Kuhn,[11] Chad Nusbaum,[1] John S. Schieffelin,[9] Bruce W. Birren,[1] Marc Forget,[12] Stuart T. Nichol,[5] Gustavo F. Palacios,[8] Daouda Ndiaye,[13] Christian Happi,[14] Sahr M. Gevao,[15] Mohamed A. Vandi,[16] Brima Kargbo,[16] Edward C. Holmes,[17] Trevor Bedford,[18] Andreas Gnirke,[1] Ute Ströher,[5,22] Andrew Rambaut,[2,19,20,22,*] Robert F. Garry,[9,22] and Pardis C. Sabeti[1,3,22,*]

[1]Broad Institute of Harvard and MIT, 75 Ames Street, Cambridge, MA 02142, USA
[2]Institute of Evolutionary Biology, Ashworth Laboratories, University of Edinburgh, Edinburgh EH9 3FL, UK
[3]Harvard University, 52 Oxford Street, Cambridge, MA 02138, USA
[4]Kenema Government Hospital, Kenema, Sierra Leone
[5]National Center for Emerging and Zoonotic Infectious Diseases and National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road NE, Mailstop-G14, Atlanta, GA 30333, USA
[6]Scripps Translational Science Institute, The Scripps Research Institute, 3344 N Torrey Pines Court, La Jolla, CA 92037, USA
[7]Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
[8]US Army Medical Research Institute of Infectious Diseases, 1425 Porter Street, Fort Detrick, Frederick, MD 21702, USA
[9]Tulane University, 1430 Tulane Avenue, SL-38, New Orleans, LA 70112, USA
[10]DNAnexus, 1975 West El Camino Real, Suite 101, Mountain View, CA 94040, USA
[11]Integrated Research Facility at Fort Detrick, Division of Clinical Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, B-8200 Research Plaza, Fort Detrick, Frederick, MD 21702, USA
[12]Médecins Sans Frontières, Rue de l'Arbre Bénit 46, 1050 Bruxelles, Belgium
[13]Université Cheikh Anta Diop, BP 5005, Dakar, Sénégal
[14]Redeemers University Nigeria, KM 46 Lagos-Ibadan Expressway, Redemption City, Ogun State, Nigeria
[15]University of Sierra Leone, A.J. Momoh St, Tower Hill, Freetown, Sierra Leone
[16]Sierra Leone Ministry of Health and Sanitation, Youyi Building, Freetown, Sierra Leone
[17]University of Sydney, Johns Hopkins Drive, Camperdown NSW 2050, Australia
[18]Fred Hutchinson Cancer Research Center, 110 Fairview Avenue North, Seattle, WA 98109, USA
[19]Centre for Immunology, Infection and Evolution, University of Edinburgh, Ashworth Laboratories, Edinburgh EH9 3FL, UK
[20]Fogarty International Center, National Institutes of Health, 31 Center Drive, MSC 2220 Bethesda, MD 20892, USA
[21]Co-first author
[22]Co-senior author
*Correspondence: dpark@broadinstitute.org (D.J.P.), a.rambaut@ed.ac.uk (A.R.), pardis@broadinstitute.org (P.C.S.)

http://dx.doi.org/10.1016/j.cell.2015.06.007

SUMMARY

The 2013–2015 Ebola virus disease (EVD) epidemic is caused by the Makona variant of Ebola virus (EBOV). Early in the epidemic, genome sequencing provided insights into virus evolution and transmission and offered important information for outbreak response. Here, we analyze sequences from 232 patients sampled over 7 months in Sierra Leone, along with 86 previously released genomes from earlier in the epidemic. We confirm sustained human-to-human transmission within Sierra Leone and find no evidence for import or export of EBOV across national borders after its initial introduction. Using high-depth replicate sequencing, we observe both host-to-host transmission and recurrent emergence of intrahost genetic variants. We trace the increasing impact of purifying selection in suppressing the accumulation of nonsynonymous mutations over time. Finally, we note changes in the mucin-like domain of EBOV glycoprotein that merit further investigation. These findings clarify the movement of EBOV within the region and describe viral evolution during prolonged human-to-human transmission.

INTRODUCTION

The 2013–2015 Western African Ebola virus disease (EVD) epidemic, caused by the Ebola virus (EBOV) Makona variant (Kuhn et al., 2014), is the largest EVD outbreak to date, with 26,648 cases and 11,017 deaths documented as of May 8, 2015 (WHO, 2015). The outbreak, first declared in March 2014 in Guinea and traced back to the end of 2013 (Baize et al., 2014), has also devastated the neighboring countries of Sierra Leone and Liberia, with additional cases scattered across the globe. Never before has an EBOV variant been transmitted among humans for such a sustained period of time.

Published EBOV Makona genomes from clinical samples obtained early in the outbreak in Guinea (three patients) and Sierra Leone (78 patients) (Baize et al., 2014; Gire et al., 2014) demonstrated that near-real-time sequencing could provide valuable information to researchers involved in the global outbreak response. Analysis of these genomes revealed that the outbreak likely originated from a single introduction into the human population in Guinea at the end of 2013 and was then sustained exclusively by human-to-human transmissions. Genomic sequencing further allowed the identification of numerous mutations emerging in the EBOV Makona genome over time. As a consequence, the evolutionary rate of the Makona variant over the time span of the early phase of the outbreak could be estimated and predictions made about the potential of this new EBOV variant to escape current candidate vaccines, therapeutics, and diagnostics (Kugelman et al., 2015a).

While the insights gleaned from sequencing early in the outbreak informed public health efforts (Alizon et al., 2014; Stadler et al., 2014; Volz and Pond, 2014), the continued human-to-human spread of the virus raises questions about ongoing evolution and transmission of EBOV. Our laboratory teams in Sierra Leone, at Kenema (Kenema Government Hospital [KGH]) and at Bo (US Centers for Disease Control and Prevention [CDC]), continued to perform active diagnosis and surveillance in Sierra Leone following our initial study (Gire et al., 2014). After a 6-month delay of sample shipment due to regulatory uncertainty about inactivation protocols, we again began to determine EBOV genome sequences. We have sequenced samples at high depth and with technical replicates to characterize genetic diversity of EBOV both within (intrahost) and between (interhost) individuals. To support global outbreak termination efforts, we publicly released these genomes prior to publication as they were generated, starting with a first set of 45 sequences in December 2014 and continuing with regular releases of hundreds of sequences through May 2015.

Here, we provide an analysis of 232 new, coding-complete EBOV Makona genomes from Sierra Leone. We compared these genomes to 86 previously available genomes: 78 unique genomes from Sierra Leone (Gire et al., 2014), 3 genomes from Guinea (Baize et al., 2014), and 5 from healthcare workers infected in Sierra Leone and treated in Europe. We use this combined data set obtained from 318 EVD patients during the height of the epidemic in Sierra Leone and Guinea to better understand EBOV transmission within Sierra Leone and between countries. In addition, we use it to understand viral population dynamics within individual hosts, the impact of natural selection, and the characteristics of the now hundreds of new mutations that have emerged over the longer course of the epidemic.

RESULTS

232 New Ebola Virus Makona Genomes from Sierra Leone

We performed massively parallel genome sequencing on 673 samples from two EVD patient cohorts. The first cohort included 575 blood samples from 484 EVD patients confirmed by laboratory staff at KGH from June 16 through September 28, 2014. The second cohort included blood samples from 88 EVD patients from throughout Sierra Leone confirmed at Bo by CDC laboratory staff from August 20, 2014 through January 10, 2015. Samples from both EVD cohorts were sequenced using previously described methods (Experimental Procedures; Matranga et al., 2014; Gire et al., 2014).

We implemented a new computational pipeline, viral-ngs:v1.0.0, for viral genomic de novo assembly, intrahost variant calling, and genome analysis and annotation. This pipeline is available via open-source software (Park et al., 2015) and utilizes a generalized workflow engine to run on a wide variety of computer hardware configurations (Köster and Rahmann, 2012). Through a partnership with DNAnexus, this pipeline is also available in a secure cloud-compute environment to enable consistent analyses across laboratories with limited computational resources (Experimental Procedures).

Using this pipeline, we successfully assembled 232 EBOV Makona coding-complete genomes (150 from KGH and 82 from the CDC cohort, spanning June 16 to December 26, 2014). Each assembled sequence was at least 18.5 kb in length, with a maximum of 6% ambiguous base calls per genome. The median assembly had 374× coverage, was 18.9 kb long, and had no ambiguous bases. Despite extensive sequencing, successful full-genome assembly was difficult to obtain from the KGH cohort (73% failed genome assemblies; 374× mean coverage; Table S1), compared to a previous cohort from the same laboratory, described in Gire et al. (2014) (11% failed genome assemblies; 2,000× mean coverage). The high assembly failure rate of the more recent KGH cohort is likely due to the mandatory in-country implementation of a new EBOV sample deactivation protocol and to long delays for sample shipments amidst the outbreak response (see Experimental Procedures). In contrast, only 7% of samples from the CDC cohort failed to assemble. However, these samples had been pre-selected for sequencing based on high EBOV titers, as estimated by qPCR. In addition, the CDC cohort samples were collected more recently, did not remain in lysis buffer for an extended period, and were subjected to a different sample deactivation protocol than the KGH cohort samples.

While we are continuing attempts to glean genomic information from compromised samples of the recent KGH cohort, important information may have been lost. In particular, samples from many EBOV-infected health-care workers at KGH, which could provide important insights into hospital-based transmissions, were compromised.

In combination with the 86 previously published EBOV Makona genomes (Gire et al., 2014), we analyzed a total of 318 genomes (see Experimental Procedures), all aligned against the earliest sampled Guinean genome (GenBank: KJ660346.2). In this set, we observed 464 single-nucleotide polymorphisms (SNPs; 125 nonsynonymous, 176 synonymous, and 163 noncoding). We also observed five single-base insertions and two double-base insertions in noncoding regions. We mapped all of the variants to primer-binding sites for known sequence-based diagnostics (Kugelman et al., 2015a) and found no mutations in these sites that were present in more than one Sierra Leonean sample (Table S2).

We constructed a second, independent genome library for each of 150 high-quality samples from the KGH cohort to reliably determine intrahost single-nucleotide variants (iSNVs) at low frequencies (Gire et al., 2014). We identified 247 iSNVs (25 insertion/deletions that were excluded from all analyses, 73 nonsynonymous, 71 synonymous, and 78 noncoding), including 21 iSNVs shared by multiple patients.

Very recently, another 175 EBOV Makona genomes were published based on a cohort from Sierra Leone, mostly sampled from the area of Freetown in the Fall of 2014 (Tong et al., 2015). Although these data were not included in our analyses, they are unlikely to significantly alter our primary findings (Figure S1).

Limited Ebola Virus Exchange across the Sierra Leonean Border

A previous study of EBOV Makona sequences elucidated viral transmission and evolution during the early stages of the outbreak in Sierra Leone (Gire et al., 2014) from late May to early June, 2014. The first reported EVD cases in Sierra Leone stemmed from two genetically distinct EBOV Makona lineages, believed to have been introduced from Guinea. One of these lineages (SL1) was more closely related to the then-available three Guinean genomes (two to five mutations) than the second lineage (SL2), which was characterized by four additional mutations. This finding suggested that SL2 had evolved from SL1 some months before it was observed in Sierra Leone. A third lineage (SL3), derived from SL2, emerged in mid-June 2014. SL3 differs from SL2 by a single mutation at position 10,218, first found as an intrahost variant (polymorphism within one individual) at a low frequency. SL3 became the most prevalent lineage in Sierra Leone during the first 3 weeks of the outbreak there, with SL1 disappearing soon after the appearance of SL3. The SL3-defining mutation is epidemiologically important, as it is the first commonly circulating mutation observed to arise within Sierra Leone's borders.

As the epidemic developed within Sierra Leone, the SL3 lineage continued to dominate the viral population within the country, with no evidence for additional imported EBOV lineages. In our data set, 97% of the genomes carry the SL3 mutation and the remainder belong to SL2 (Figure 1A). These results link all Sierra Leonean EVD cases to the initial introduction of EBOV into Sierra Leone, and they provide further evidence that all EVD cases during this outbreak arose from human-to-human transmission rather than from further zoonotic introductions from the unknown EBOV reservoir. This means that no newly imported viral diversity was detected after the initial introduction (Gire et al., 2014); all newly sampled viruses likely descended from those sequenced in the initial weeks of the outbreak. The genetic similarity of these viruses suggests that importation from other countries was minimal, although we cannot definitively rule out a re-introduction from elsewhere for the SL2 viruses (3%) in our data set.

Similarly, publicly available EBOV genomes from this outbreak can shed light on exportation of EBOV from Sierra Leone into other countries. All published genomes from elsewhere, including 26 from Liberia and 4 from Mali, lack the Sierra Leone-defining SL3 mutation (Figure 1B and Experimental Procedures). Given that 97% of Sierra Leonean EBOV sequences have the SL3 variant, extensive exportation would result in the spread of SL3 EBOV genomes, a spread that is not seen in the limited samples available to date. At least in Sierra Leone, and with the exception of events at the onset of the epidemic, transmission has likely been primarily within national borders (Figure S2 and Experimental Procedures), rather than by free interchange with neighboring countries.

Human-to-Human Transmission of Multiple EBOV Genomes

Intrahost variants (iSNVs) that appear during the course of the epidemic may provide valuable information about human-to-human transmission. In particular, shared iSNVs have been used to estimate the relative size of the transmission bottleneck (Emmett et al., 2015) and to identify human-to-human transmission chains (Gire et al., 2014). In the current data set, which includes 85 samples with at least one iSNV (Figure S3A), several iSNVs are shared among two or more patients, often spanning several months of the EVD epidemic (Figure 2A). The existence of shared iSNVs could be explained by patient infection from multiple sources (superinfection), sample contamination, recurring mutations (with or without balancing selection to reinforce mutations), or co-transmission of slightly diverged viruses that arose by mutation earlier in the transmission chain.

We can rule out superinfection and contamination as primary explanations for the iSNVs in our data because none of the iSNVs are located at common SNP positions. For example, a SNP at position 14,019 is at intermediate frequency in the population (found in ~40% of samples we sequenced) and defines the SL4 lineage (Figure 1A). If superinfection were common among EVD patients, we would expect to sometimes see both SL3 and SL4 viruses in the same patient, which would appear as an iSNV at that position. Contamination would result in a similar pattern, with intermediate-frequency SNPs appearing as iSNVs in contaminated samples. Additionally, contamination would be most visible in low-coverage, low-RNA-content samples because contaminants would make up more of the RNA available for sequencing, whereas samples with extremely high coverage would be the most visible contaminants (Figure S3B). The highest coverage sample (G4960.1) contains genomes belonging to lineage SL3 only and lacks the SL4 SNP, so if there were widespread contamination, we would see a low-frequency iSNV at position 14,019 in SL4 samples with iSNVs. Since SL3 and SL4 samples were processed together (eight of nine sequencing batches contained multiple samples from both lineages) and we saw no instances of an iSNV at that position, we conclude that superinfection and contamination are not important contributors to iSNVs.

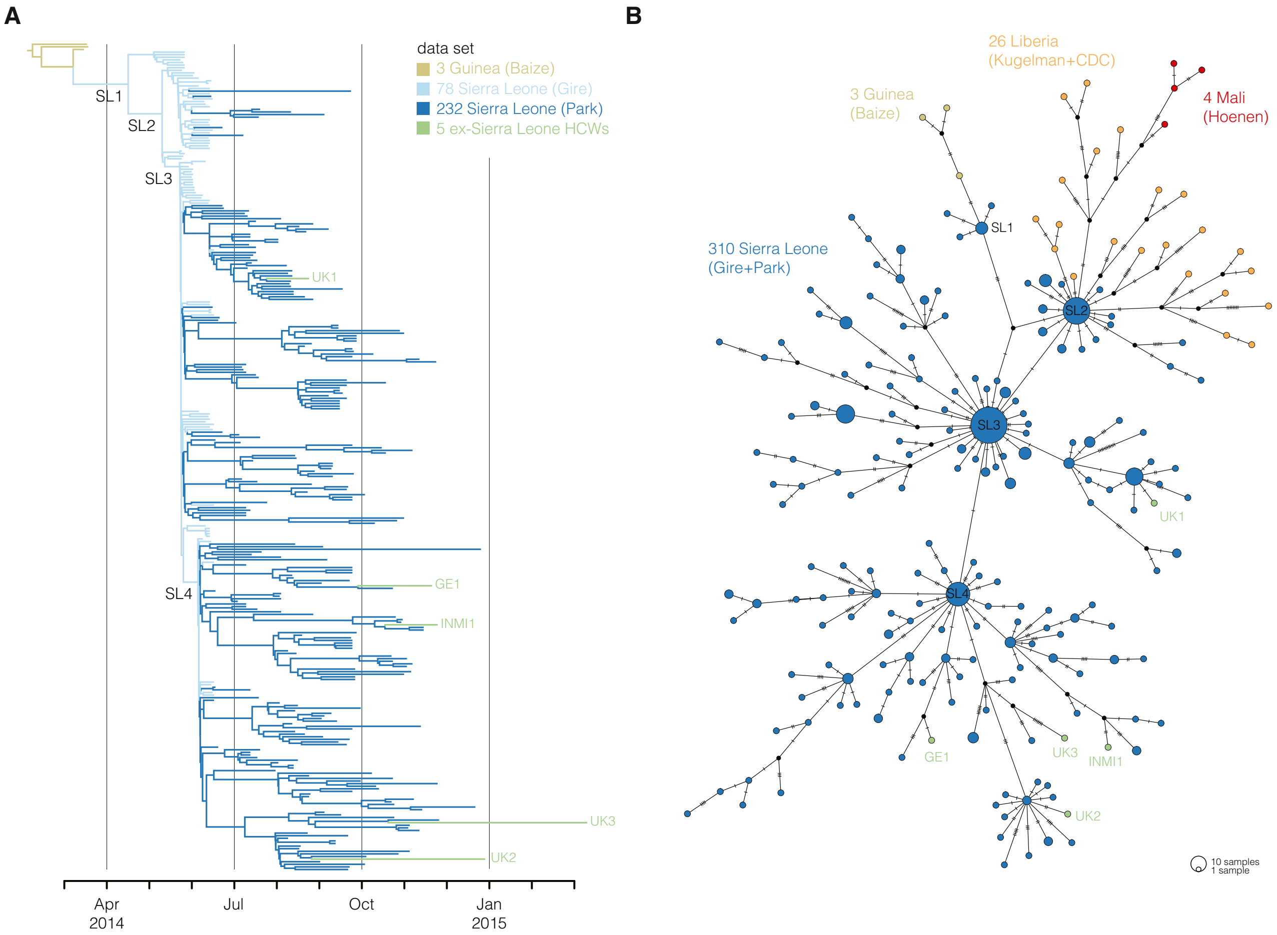


Figure 1. Within and between Country Genomic Relationships of Ebola Virus Makona

(A) Phylogenetic and temporal placement of recently sequenced Ebola virus (EBOV) within Sierra Leone. New EBOV genomes (232 genomes, dark blue), sampled from June 16 through December 26, 2014, provide a high-resolution view of the accumulated genetic diversity and fill in the missing ancestry between EBOV Makona genome data sets. The maximum clade credibility (MCC) tree was inferred using Bayesian evolutionary analysis by sampling trees (BEAST), with tips anchored to sampling date. Tips are labeled for EBOV from five non-African health-care workers (HCWs) infected in Sierra Leone and treated in Europe (sequenced by other groups, light green). Previously described nested EBOV Makona lineages SL1, SL2, and SL3 Gire et al. (2014), as well as a new lineage SL4, are labeled at their most-recent common ancestor (MRCA) nodes.

(B) Lack of EBOV Makona SL3 spread to Liberia or Mali. Shown is a median-joining haplotype network constructed from a coding-complete EBOV genome alignment including 340 EBOV Makona sequences. Each colored vertex represents a sampled viral haplotype, with colors indicating countries of origin. Colors are as in (A), with the exception that the distinction is no longer made between older (Gire) and newer (Park) Sierra Leonean data sets (both are now dark blue), and two additional countries are shown (Liberia in yellow, Mali in red). The size of the each vertex is relative to the number of sampled isolates. Hatch marks indicate the number of mutations along each edge.

See also Figures S1 and S2.

The remaining possible sources for persistently shared iSNVs are co-transmission and recurrent mutation. In either case, the iSNV could be maintained by balancing selection or could be evolving neutrally. Figure 2A suggests that selection is not the primary cause of persistence, since synonymous and nonsynonymous variants are equally common among the shared iSNVs, and selective pressures are likely to be different for the two classes of variant. All shared iSNVs are unlikely to be simply the product of recurring mutation: if they were, they should have a frequency spectrum heavily weighted toward low frequency, characteristic of new mutations. However, that is not the case. For example, the variant at position 18,911 is found at >15% frequency in eight different samples (Figure S3C), a much higher frequency than expected if the change represented a de novo mutation in each sample.

In summary, we conclude that a combination of human-to-human transmission and recurrent mutations is likely responsible for the iSNV pattern observed in Figure 2A. This hypothesis is supported by the iSNV at position 18,911: samples containing this variant often cluster on the phylogenetic tree (Figure 2B), although more isolated samples may represent separate mutation events. More generally, pairs of samples that share an iSNV are typically located near one another phylogenetically; these pairs are separated by an average of 0.16 years of

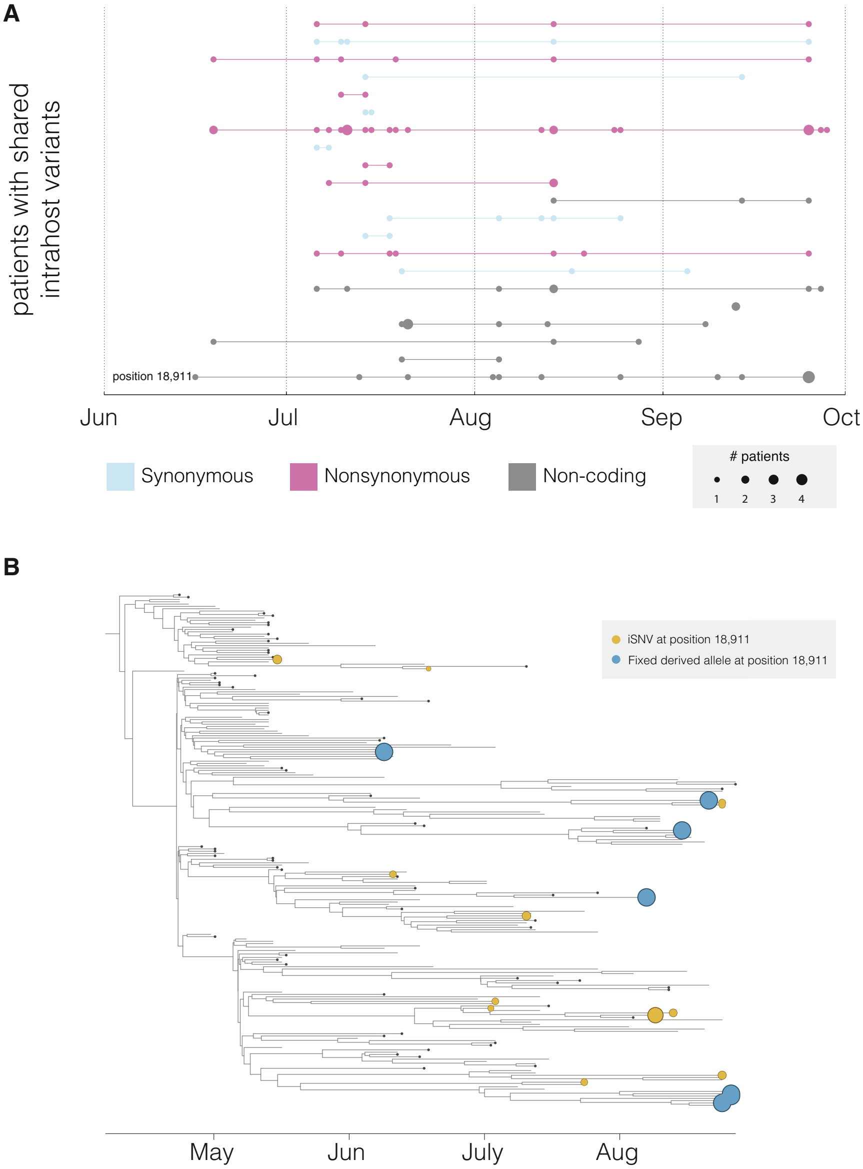


Figure 2. Evidence for Host-to-Host Transmission of Multiple Ebola Virus Makona Genomes

(A) Certain intrahost variants (iSNVs) appear in samples throughout the 2013–2015 EVD epidemic, suggesting that iSNVs can be transmitted between patients. Variants shared between two or more samples are shown as rows of connected points; each row is a genomic position (ordered by position along the genome, top to bottom), and each point indicates the presence of the iSNV in a patient.

(B) Phylogenetic placement of derived alleles at genomic position 18,911 implies both repeated transmission within clades as well as some amount of recurrent mutation. Colored tips are sized according to frequency of iSNV at position 18,911. Tips with small black points are those with iSNV calls at any position; other tips represent samples with no iSNV calls. This figure shows only the portion of the tree relevant for this analysis; large branches with no SNPs or iSNVs at position 18,911 are not shown.

See also Figure S3.

evolution, whereas random pairs are separated by an average of 0.30 years ($p < 10^{-4}$, randomization test). These results suggest transmission of iSNVs in at least some cases and therefore suggest that the transmission bottleneck is wide enough to facilitate the transmission of low- or intermediate-frequency variants between hosts.

Viral Evolution during a Prolonged EVD Epidemic

We previously reported that new mutations accumulated more rapidly in the viral population early in the outbreak than over the long-term in the reservoir (Gire et al., 2014). We hypothesized then that the higher rate early in the outbreak resulted from incomplete purifying selection—that is, we were detecting transient nonsynonymous variants that would later be removed by purifying selection (Pybus et al., 2007; Bedford et al., 2011). The observed evolutionary rate is thus not an estimate of the underlying mutation rate since some deleterious mutations are purged by selection before they can be detected. But neither is it an estimate of the long-term substitution rate since other deleterious mutations have not been eliminated by selection at the time of analysis. We hypothesized that the EBOV Makona evolutionary rate would decline following the addition of genomes covering a longer evolutionary timescale. Such a decline is well characterized in members of other species (Duchene et al., 2014; Ho et al., 2005). With the present data set, we were able to examine the evolution of the virus over a longer time period. We found that the most probable estimated evolutionary rate of EBOV Makona is indeed markedly lower (mean posterior rate = 1.25×10^{-3} substitutions per site per year) and is closer to the long-term rate than to the rate estimated early in the outbreak (Figures 3A and S4).

How purifying selection acts at different timescales can also be seen in the distribution of mutations in the EBOV Makona genealogy. Deleterious mutations are more likely to result in transmission-impaired viruses and dead-end infections and may therefore only be present in individual patients. Mutations unique to individual patients are those that occur on the external branches of the phylogenetic tree, whereas internal branch mutations are those present in multiple samples in our data set. Thus, in the model of incomplete purifying selection, we expect external branches to be characterized by a higher rate of nonsynonymous substitution than internal branches; in the latter, selection has had more opportunity to filter out deleterious mutants. Internal branches, by definition, have produced multiple

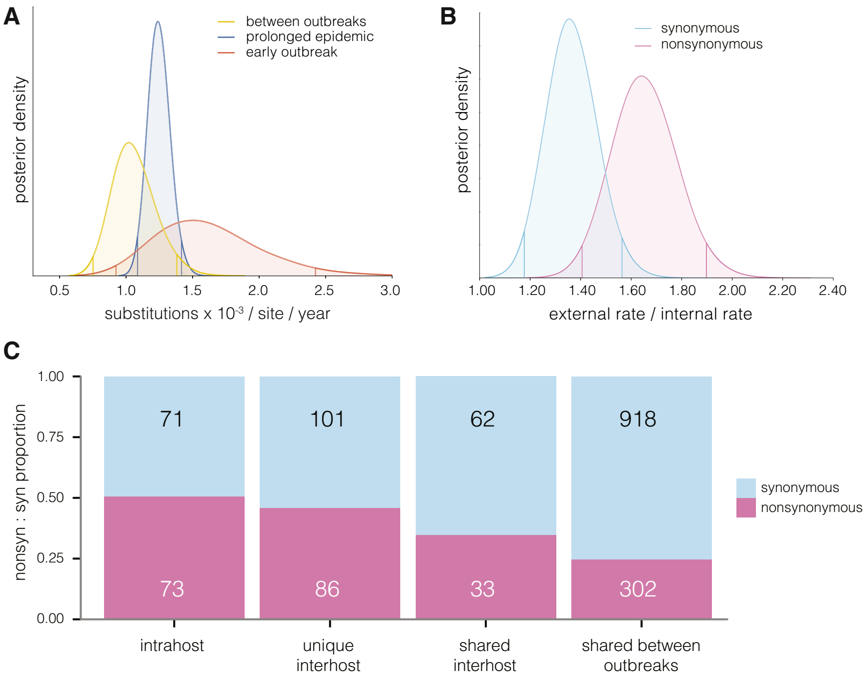


Figure 3. Ebola Virus Evolution during a Prolonged EVD Epidemic

(A) Estimates of EBOV evolutionary rates at three timescales: decades (yellow, all known EVD outbreaks), months (blue, Baize + Gire + Park), and weeks (red: Baize + Gire).

(B) Purifying selection. We estimated nonsynonymous (red) and synonymous (blue) substitution rates on external (unique to an isolate, potential dead end) and internal (shared by multiple isolates, evidence of human-to-human transmission) branches. Nonsynonymous mutations accumulate faster on external branches than on internal branches. For synonymous mutations, the difference between external and internal branches is less pronounced.

(C) Enrichment for nonsynonymous mutations at shorter timescales. Intrahost (all variants that appear within a single host at less than 100% frequency); unique interhost (SNPs fixed in exactly one individual); shared interhost (SNPs fixed in two or more individuals); shared between EVD outbreaks (internal branch SNPs on a between-outbreak tree).

See also Figure S4.

descendent lineages and are thus less likely to include mutations with fitness costs. To test this hypothesis, we estimated the numbers of nonsynonymous and synonymous changes on the virus genealogy and recovered their accumulation rates (Figure 3B). Nonsynonymous mutations indeed occurred at lower frequency on internal than on external branches, suggesting that most are removed by purifying selection because of their fitness costs and hence represent evolutionary dead ends. Synonymous mutations, which likely have less impact on fitness, occurred at more comparable frequencies on internal and external branches.

The relationship between the effectiveness of purifying selection and its duration is also apparent in the overall pattern of nonsynonymous mutations in our data set. Selection filters the accumulation of coding variants in the EBOV genome (Figures 3C and 4A). Nonsynonymous mutations, which are more likely to be deleterious, make up a decreasing fraction of coding mutations as we analyze longer timescales: intrahost variants > individual patients (external branches) > multiple patients (internal branches) > between outbreaks. The fraction seen between outbreaks represents the effect of long periods of evolution in the unknown EBOV reservoir. As selection acts to remove deleterious alleles over time, fewer nonsynonymous mutations can be detected. This pattern holds true across the EBOV Makona genome (Figure 4A).

Possible Host Effects on the Viral Genome

Although we observe less constraint on nonsynonymous changes during the 2013–2015 epidemic than between outbreaks, one anomaly is the genomic sequence encoding the mucin-like domain of the EBOV glycoprotein (GP), for which we observe more nonsynonymous substitutions than expected under neutrality, both within and between EVD outbreaks. Selective pressure acting on a region can be estimated with the standard statistic d_N/d_S, which has an expected value of 1.0 for neutral evolution and less than 1 for purifying selection; in the mucin-like domain, the mean posterior d_N/d_S within this outbreak is 4.74, and between outbreaks is 1.44 (Figure 4A). GP is the only surface-exposed viral protein on EBOV virions, and as such, it is the primary target of antibodies (Murin et al., 2014). This finding therefore raises the possibility that antibodies might be driving diversifying selection and rapid evolution in this region. This observation is based on a very small number of substitutions (eight nonsynonymous and four synonymous within the outbreak), however, and is not statistically significant (posterior probability that d_N/d_S is elevated within-outbreak = 92.9%); the situation should be clarified as more sequencing becomes available. If diversifying selection is occurring here, then the observed changes are very unlikely to represent population-level selection for transmission among humans; this would only occur if previously infected individuals were frequently being exposed to new infections. Instead, we hypothesize that these changes represent within-host selection for EBOV to escape a developing humoral immune response.

To test the hypothesis that antibodies drive diversifying selection of GP, we looked for enrichment of mutations within B cell epitopes within that protein. Effective humoral immunity depends on antibody binding to specific B cell epitopes (Becquart et al., 2014; Murin et al., 2014). Using experimentally determined B cell epitopes obtained from the Virus Pathogen Database and Analysis Resource (ViPR; Pickett et al., 2012), we found that nonsynonymous mutations in GP do indeed occur more frequently in epitopes than expected by chance (Figure 4B). This correlation supports the hypothesis that humoral immunity exerts selective pressure on the virus, driving immune evasion via accumulation of nonsynonymous mutations within GP B cell epitopes.

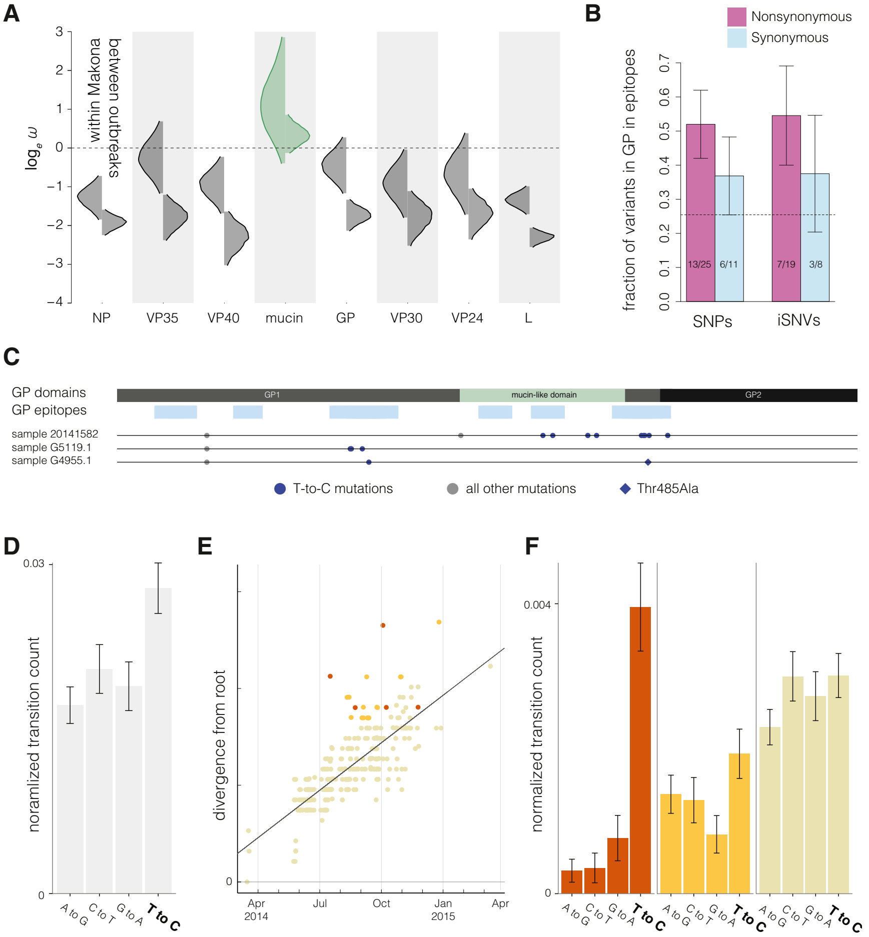


Figure 4. Evidence for Host Effects on Ebola Virus Makona Evolution

(A) Nonsynonymous variants are enriched in the mucin-like domain of GP. Estimates of log(ω) (a.k.a., log(d_N/d_S)) per coding sequence within the Western African EVD outbreak (left) and between EVD outbreaks (right) demonstrate gene-specific patterns of natural selection.

(B) Nonsynonymous variants are enriched in B cell epitopes of GP. We calculated the fractions of nonsynonymous (NS) and synonymous (S) consensus SNPs and intrahost variants (iSNVs) within experimentally determined B cell epitopes (data from ViPR; Pickett et al., 2012). Dotted line represents the fraction of GP amino acids in ViPR epitopes. Nonsynonymous SNPs (p = 0.004) and iSNVs (p = 0.037) in GP occur more frequently in epitopes than expected by chance (two-sided exact binomial test). Numbers indicate fraction of each variant type within GP epitope regions. Error bars represent binomial sampling intervals.

(C) Local enrichment of T-to-C mutations within GP B cell epitopes. We observed five sequences with short stretches (<200 nucleotides) of concentrated T-to-C mutations. Of these five sequences, two (shown here, samples 20141582 and G5119.1) contain stretches of T-to-C SNPs (blue points) within GP epitopes (light blue bars). Additionally, we observe a T-to-C mutation at amino acid position 485 (blue diamond) in three samples (one shown here, G4955.1), which is otherwise completely conserved among members of all ebolavirus species (Olal et al., 2012).

(D) Genome-wide increase in T-to-C mutations. We observe more T-to-C transitions within the 2013–2015 outbreak than any other transition, after correcting for nucleotide content. Error bars represent binomial sampling intervals.

(E and F) Elevated T-to-C rates are genome wide but are limited to a subset of sequences. Accumulation of mutation increases linearly with time. However, some individual samples show more genetic distance than expected based on sample date. Samples with short stretches of T-to-C mutations (orange) show a significant enrichment of T-to-C mutations, as expected. Excluding these samples, the top 5% of samples by genetic distance (yellow) lack localized stretches but still show moderate enrichment of T-to-C mutations genome wide. The bottom 95% of samples (beige) show no enrichment of T-to-C mutations. Error bars represent binomial sampling intervals.

Visual inspection identified a subset of sequences that are more likely to contain B cell escape variants (Figure 4C). In particular, three sequences (e.g., G4955.1) had a threonine-to-alanine mutation at GP amino acid position 485, a conserved threonine that is required for in vivo protection by the 14G7 antibody (Olal et al., 2012). Additionally, two sequences had short stretches of T-to-C mutations in GP (four or more T-to-C mutations within a 200 nucleotide region; Figure 4C), both of which occur within B cell epitopes.

Similar patterns of excess T-to-C mutations within short regions were also observed by Tong et al. (2015). In our data set of 318 genomes, five possessed obvious stretches of T-to-C mutations within short regions. We also tested more broadly whether excessive T-to-C mutations occurred in all sequences and found a significant enrichment of T-to-C transitions relative to all other types of transitions (Figure 4D). To determine whether viral sequence divergence is related to T-to-C transition enrichment, we compared relative T-to-C transition rates in sequences with stretches of T-to-C mutations (n = 5) to the top 5% of remaining sequences by sequence divergence (n = 15) and to the bottom 95% of sequences (n = 298) (Figure 4E). While the sequences with T-to-C stretches showed the strongest T-to-C enrichment, we found moderate enrichment of T-to-C transitions in the 5% most divergent sequences.

DISCUSSION

Our findings from 232 EBOV Makona genomes sampled in Sierra Leone over 7 months during the 2013–2015 EVD outbreak in Western Africa demonstrate the value of continued sequencing throughout an epidemic. We tracked the movement of EBOV throughout Sierra Leone and determined the frequency of EBOV movement into and out of that country. Although it is not unlikely that the virus continued to cross the national borders of Sierra Leone throughout the epidemic, these observations suggest that, at least in late 2014, cross-border introductions were not an important factor in the development of the epidemic. We were unable, however, to draw any conclusions about export to Guinea since few EBOV sequences from there are currently available.

The sequence data display EBOV Makona evolution in the context of prolonged human-to-human transmission and provide an updated view of genomic diversity. Based on the rates of nonsynonymous and synonymous changes that are shared or are unique to an individual host, we concluded that purifying selection becomes increasingly effective over time, as it has more opportunity to remove deleterious mutants.

While the effects of purifying selection in this extended EVD outbreak are clear, these evolutionary changes do not imply that positive selection or adaptation to humans are occurring. Rather, the data suggest that evolutionary changes over time through natural selection are sufficient to remove newly arisen alleles that are less fit in the human environment. To date, no published study has found experimental evidence of selection for alleles beneficial to the virus within the current outbreak.

It is important to recognize, however, that the long-term human-to-human transmission observed during the 2013–2015 EVD outbreak is historically unique for EBOV. At the beginning of each EVD outbreak, EBOV enters the human population with little or no genetic diversity. In the case of the current EVD outbreak, EBOV has now maintained fitness while expanding across a much larger space of genetic diversity than in previous EVD outbreaks, the largest of which comprised only 318 human infections. This degree of diversity will undoubtedly affect researchers' ongoing efforts to develop or improve candidate diagnostics, vaccines, and therapeutics for EVD, many of which are targeting EBOV sequences directly (PCR, nucleic-acid based therapeutics) or indirectly (antibody cocktails).

The mucin-like domain of the EBOV glycoprotein, in contrast to the rest of the EBOV genome, appeared to be under diversifying selection based on a high ratio of nonsynonymous-to-synonymous mutations. While not statistically significant because of the small number of SNPs in the region, our observation is in agreement with many previous studies (Sanchez et al., 1998; Wertheim and Worobey, 2009). As the EBOV GP, especially the mucin-like domain, is the target of many antibodies, a plausible hypothesis is that the humoral immune response exerts selective pressure on GP, resulting in an accumulation of nonsynonymous mutations. In support of this hypothesis, regions of GP corresponding to experimentally determined B cell epitopes are significantly enriched in nonsynonymous, but not in synonymous, variants. There are two important caveats to this analysis: (1) these epitopes are determined in vitro and therefore may not be epitopes in vivo if they are not immunodominant, and (2) there is no experimental evidence to suggest that the majority of observed variants disrupt antibody binding to these epitopes.

While further experimental testing is required to validate an immune evasion hypothesis, we have highlighted a few prime candidates to consider. Genomes from three samples share a threonine-to-alanine mutation at GP amino acid position 485, a position that is conserved among all members of the Ebolavirus genus. This position is indispensable for binding of the protective antibody 14G7 (Olal et al., 2012); the observed variant at this site may therefore be the result of escape from antibody-mediated selection. Additionally, two samples each possess multiple mutations within a single experimental B cell epitope in GP, which are likely to evade antibody recognition if those regions are relevant epitopes in vivo.

Intriguingly, the two samples with multiple mutations within a single B cell epitope each possess a distinct short stretch littered with T-to-C transitions, a phenomenon also observed in Tong et al. (2015). Excessive T-to-C and A-to-G mutation of virus genomes has been observed previously as a result of adenosine deaminases acting on RNA (ADARs; Gélinas et al., 2011; Zahn et al., 2007; Carpenter et al., 2009). When acting on viral genomic RNA, ADARs cause a pattern of excess A-to-G transitions that are represented by T-to-C transitions in our data set. These transitions are known to occur either promiscuously within 200 nucleotide stretches or in a sequence-specific manner; therefore, we investigated both possibilities. While only five of the 318 sequences in our data set contained obvious T-to-C stretches, we showed that the top 5% of sequences by sequence divergence, excluding the five sequences with T-to-C stretches, were also moderately enriched for T-to-C transitions across the genome. The remaining 95% of sequences appeared to show no enrichment. We do not know whether this phenomenon is caused by ADAR acting upon genomic RNA, as we cannot exclude the possibility of bias by the EBOV RNA polymerase or other effects. Additionally, it is yet unclear whether these T-to-C mutations have an anti-viral or other effect on viral fitness. These questions open avenues of research into molecular mechanisms shaping EBOV evolution.

The results of some of the specific genome analysis methods that we introduced here, while promising, will require denser EBOV genome sampling to yield sufficient information to influence the EVD outbreak response. Among these methods is transmission analysis, which could prove valuable for improved understanding of hospital-based transmissions and therefore for improved infection control. Inference of the ancestral genetic state is often straightforward, with clear patterns of new variations layering on previously existing variations; viruses that appear to be descended from others in the same data set are separated only by new mutations that are seen nowhere else in the data set. This kind of genetic relationship does not guarantee a transmission relationship between two patients since many viruses can share identical genomes. However, since viruses with identical genomes are often epidemiologically related (Gire et al., 2014), we can infer that viruses that appear to descend from other viruses in our data set are either in or epidemiologically close to the same transmission chain.

Unfortunately, long delays of shipping samples from the field and required changes to the EBOV inactivation protocol caused severe degradation of many samples, which prevented identification of variants and transmission analysis. This loss should serve as a reminder that standardized and optimized protocols for sample collection, virus deactivation, and shipment are crucial for a rapid worldwide response to any new infectious disease outbreak. An important future research effort will be aimed at understanding which certified EVD sample deactivation protocols are best suited for high-quality genomic sequencing. Complications with sample shipment also emphasize the need for establishing in-country sequencing capabilities either before or at the onset of future EVD outbreaks (Folarin et al., 2014).

Beyond coordinated field and experimental responses, a culture of rapid data sharing is critical for teams around the world to have the best current information about a circulating virus or ongoing disease (Yozwiak et al., 2015). In light of this need, we released all data discussed in this paper publicly as they were generated, beginning in December 2014, well in advance of our own analysis. We have previously described our high-depth sequencing protocols (Matranga et al., 2014), and we are also now making available our computational analysis pipeline, in the hope that they will assist the many laboratories engaged in viral genomic research. As more EBOV genomic data become available, in particular for poorly covered Liberia and Guinea, the scientific community can together obtain a broader picture of transmission and evolution of EBOV Makona during the EVD epidemic.

EXPERIMENTAL PROCEDURES

Sample Preparation from Kenema Government Hospital

This study included 575 blood samples from 84 patients with confirmed EVD from June 16 through September 28, 2014 by KGH laboratory staff. Clinical samples were inactivated using QIAGEN AVL and ethanol in the KGH laboratory prior to shipping out of the country.

Sample Preparation from CDC Bo Laboratory

This study included 98 blood samples from 98 patients with confirmed EVD from August 20, 2014 through January 10, 2015 by CDC laboratory staff stationed in Bo, Sierra Leone. Clinical specimens from the CDC Bo laboratory in Sierra Leone were shipped to and stored at the Viral Special Pathogens Branch BSL-4 laboratory at the CDC in Atlanta, GA. Samples were inactivated and RNA was extracted using the MagMAX Pathogen RNA/DNA isolation kit (Invitrogen) and BeadRetriever (Invitrogen). Non-infectious RNA was treated with DNase I RNase-free (Roche) prior to shipment to the Broad Institute.

High-Throughput Sequencing

Host ribosomal and carrier poly(rA) RNA depletion, randomly primed cDNA synthesis, Nextera XT library construction, and 101-bp paired-end Illumina sequencing were performed as described previously (Gire et al., 2014; Matranga et al., 2014).

Ebola Virus Makona Genome Assembly and Analysis

EBOV Makona genomes were assembled from high-throughput sequencing data using an updated bioinformatics pipeline based on our previously described methods (Gire et al., 2014; Matranga et al., 2014). Of the collected samples, 150 KGH and 82 CDC samples had sufficient EBOV genome sequencing coverage for high-quality de novo genome assembly. Further description of the pipeline can be found in the Supplemental Experimental Procedures.

Our Linux-based software pipeline is publicly available at https://github.com/broadinstitute/viral-ngs (Park et al., 2015). This pipeline includes command-line tools for each of the above steps and optional Snakemake workflows (Köster and Rahmann, 2012) to automate them either sequentially or in parallel.

The assembly pipeline is also available via the DNAnexus cloud platform. RNA paired-end reads from either HiSeq or MiSeq instruments (Illumina) can be securely uploaded in FASTQ or BAM format and processed through the pipeline using graphical and command-line interfaces. Instructions for the cloud analysis pipeline are available at https://github.com/dnanexus/viral-ngs/wiki.

Genomic Epidemiology of Ebola Virus Makona

The following publicly available EBOV Makona genomes from outside of Sierra Leone do not carry the SL3-derived allele at position 10,218: 26 available genomes from Liberia (25 from Kugelman et al., 2015b, one from GenBank: KP178538.1), and all four available genomes from Mali (Hoenen et al., 2015). A median-joining haplotype network was constructed in PopART version 1.7.2 (http://popart.otago.ac.nz). Due to the presence of missing data, 1,492 sites (7.9% of total genome) were excluded from the analysis; these sites included 61 sites with variability among isolates (10.9% of all variable sites).

To reconstruct the EBOV Makona transmission history within Sierra Leone, we grouped samples into sets of one or more genetically identical viruses based on their consensus sequences. We then identified relationships between these groups, progressing from the Guinean reference genome (KJ660346.2) and ending with nine viruses sampled in Freetown (eight from our KGH and CDC cohorts and one sequenced in Italy).

Intrahost Variant Analysis

Full details of the identification and calling of intrahost variants (iSNVs) are available in the Supplemental Experimental Procedures; iSNV calls and analyses are available in Data S1. Evolutionary distances between pairs of phylogeny tips were computed from the posterior sample of trees produced by Bayesian evolutionary analysis by sampling trees (BEAST) (Drummond et al., 2012) analysis. This calculation integrates across phylogenetic uncertainty and produces a temporal evolutionary distance between phylogeny tips. We used this distance matrix to calculate the average distance between pairs of phylogeny tips that share an iSNV and compared the result to the average distance between random pairs of tips. We calculated a p value for the observed average distance by conducting a randomization test. In each random replicate, we sampled the same distribution of iSNV possessing tips as observed in the empirical data and calculated the average distance between these pairs of tips. We calculated a p value by comparing the empirical mean distance to the mean distances observed over 10,000 random replicates.

GP B Cell Epitope Analysis

Data were obtained from the NIAID Virus Pathogen Database and Analysis Resource (ViPR) online through the web site at http://www.viprbrc.org (Pickett et al., 2012). As most of the epitopes in the database are based on the Mayinga reference strain, we mapped all B cell epitopes against the Guinean reference strain (GenBank: KJ660346.2) and removed all epitopes that no longer matched perfectly, leaving 40 B cell epitopes. Overlapping epitopes were merged, and nonsynonymous and synonymous SNPs and iSNVs were scored as within or outside of epitope regions. Significance was determined by two-tailed binomial test with $\alpha = 0.05$, with the null hypothesis that variants would occur in epitope regions of GP by chance with probability 172/676, which is the fraction of GP residues GP within B cell epitopes.

Molecular Evolution

Three data sets were constructed to represent three timescales of genetic surveillance of EBOV Makona. For surveillance between EVD outbreaks, 63 publicly available sequences represent the diversity of EBOV sampled over long periods of time; these sequences include the first recorded 1976 EVD outbreak and other EVD outbreaks and exclude one outbreak occurring in the Democratic Republic of the Congo in 2014. We also included EBOV genome fragment sequences from possibly infected great ape carcasses and frugivorous bats. Fourteen sequences from Western Africa were chosen to represent the current 2013–2015 EVD outbreak. For surveillance of the early outbreak, 81 sequences (Baize et al., 2014; Gire et al., 2014) were reanalyzed, representing the earliest epidemiologically relevant and publicly available sequences. For surveillance of the prolonged epidemic, 232 EBOV genomes reported here were combined with five sequences from repatriated healthcare workers (UK1, UK2, UK3, INMI1, GE1) and the 81 sequences from the early outbreak data set.

Analyses of rates, phylogenies, and evolution were performed on all three data sets in BEAST (Drummond et al., 2012). Full details on the models and parameters are available in the Supplemental Experimental Procedures. All BEAST inputs, outputs, and analysis scripts are available in Data S2.

ACCESSION NUMBERS

Genome assemblies, annotations, and raw reads are available at NCBI on GenBank and SRA using the following BioProject IDs: PRJNA257197 (samples from Kenema Government Hospital) and PRJNA283385 (samples from CDC Bo Lab). Note that PRJNA257197 also includes all previously published data from Gire et al. (2014).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, two tables, and two data files and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2015.06.007.

AUTHOR CONTRIBUTIONS

The contributions of each author are too extensive to list in detail. But among the first five and last four authors, A.G. and R.F.G. collected samples. A.G. and S.L.M.W. processed samples for sequencing. D.J.P, G.D., S.W., S.L.M.W, and A.R. analyzed sequence data. D.J.P, G.D., S.W., S.L.M.W., U.S., A.R., and P.C.S. wrote the paper. U.S., A.R., R.F.G., and P.C.S. jointly supervised this work.

ACKNOWLEDGMENTS

We thank KGH staff who died of EVD (including M. Fonnie, A. Moigboi, A. Kovoma, M. Fullah, and S.H. Khan), the Office of the President of Sierra Leone (President E. Koroma, M. Jones), the Sierra Leone Ministry of Health and Sanitation, the Kenema District Health Management team, and the Kenema Lassa fever program for their immense efforts in the EVD outbreak response. We thank Public Health England (UK1, UK2, UK3), IRCCS Lazzaro Spallanzani (INMI1), and the University of Geneva (GE1) for providing EBOV genome sequences from samples of EVD patients exported from Sierra Leone. We thank the drivers, pilots, phlebotomists, non-governmental organizations, district medical officers, and district surveillance officers for their help with sample collection and logistics in Sierra Leone. We want to especially thank the Médecins Sans Frontières (MSF) operation centers for their continuing support of the US Centers for Disease Control and Prevention (CDC) laboratory in Bo, Sierra Leone and the World Health Organization (WHO) for their support of the preceding CDC laboratory operation in Kenema, Sierra Leone.

This work was supported by European Union grant FP7/2007-2013 278433-PREDEMICS and European Research Council grant 260864 (A.R.); Natural Environment Research Council grant D76739X (G.D.); NIH U54 GM111274 (T.B.); NIH grant GM080177 (S.W.); NIH grant 1U01HG007480-01 (C.H.); National Science Foundation Graduate Research Fellowship Grant No. DGE 1144152 (A.E.L.); the National Health and Medical Research Council, Australia (E.C.H.); the Defense Threat Reduction Agency (USAMRIID); NIH/NIAID U19AI110818 (Broad Institute); the Bill and Melinda Gates Foundation OPP1123407 (Broad Institute); and NIAID HHSN272200900049C (Harvard/Tulane). This work was funded, in part, through Battelle Memorial Institute's prime contract with the US National Institute of Allergy and Infectious Diseases (NIAID) under contract number HHSN272200700016I. Subcontractors to Battelle Memorial Institute who performed this work are: J.H.K., an employee of Tunnell Government Services, Inc. R.F.G. is co-founder of Zalgen Labs.

The Virus Pathogen Database and Analysis Resource (ViPR) has been wholly funded with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under contract number HHSN272201400028C.

This paper was authored in Authorea, and its edit history is available here: https://www.authorea.com/users/10734/articles/19957.

The content of this publication does not necessarily reflect the views or policies of the US Department of Health and Human Services (Centers for Disease Control and Prevention, National Institutes of Health) or the US Army.

Received: May 13, 2015
Revised: May 26, 2015
Accepted: June 1, 2015
Published: June 18, 2015

REFERENCES

Alizon, S., Lion, S., Murall, C.L., and Abbate, J.L. (2014). Quantifying the epidemic spread of Ebola virus (EBOV) in Sierra Leone using phylodynamics. Virulence *5*, 825–827.

Baize, S., Pannetier, D., Oestereich, L., Rieger, T., Koivogui, L., Magassouba, N., Soropogui, B., Sow, M.S., Keïta, S., De Clerck, H., et al. (2014). Emergence of Zaire Ebola virus disease in Guinea. N. Engl. J. Med. *371*, 1418–1425.

Becquart, P., Mahlakõiv, T., Nkoghe, D., and Leroy, E.M. (2014). Identification of continuous human B-cell epitopes in the VP35, VP40, nucleoprotein and glycoprotein of Ebola virus. PLoS ONE *9*, e96360.

Bedford, T., Cobey, S., and Pascual, M. (2011). Strength and tempo of selection revealed in viral gene genealogies. BMC Evol. Biol. *11*, 220.

Carpenter, J.A., Keegan, L.P., Wilfert, L., O'Connell, M.A., and Jiggins, F.M. (2009). Evidence for ADAR-induced hypermutation of the Drosophila sigma virus (*Rhabdoviridae*). BMC Genet. *10*, 75.

Drummond, A.J., Suchard, M.A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. *29*, 1969–1973.

Duchene, S., Holmes, E.C., and Ho, S.Y. (2014). Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. Proc. Biol. Sci. *281*. Published online July 7, 2014. http://dx.doi.org/10.1098/rspb.2014.0732.

Emmett, K.J., Lee, A., and Rabadan, R. (2015). High-resolution genomic surveillance of 2014 Ebolavirus using shared subclonal variants. PLoS Curr. Published online February 9, 2015. http://dx.doi.org/10.1371/currents.outbreaks.c7fd7946ba606c982668a96bcba43c90.

Folarin, O.A., Happi, A.N., and Happi, C.T. (2014). Empowering African genomics for infectious disease control. Genome Biol. *15*, 515.

Gélinas, J.-F., Clerzius, G., Shaw, E., and Gatignol, A. (2011). Enhancement of replication of RNA viruses by ADAR1 via RNA editing and inhibition of RNA-activated protein kinase. J. Virol. *85*, 8460–8466.

Gire, S.K., Goba, A., Andersen, K.G., Sealfon, R.S.G., Park, D.J., Kanneh, L., Jalloh, S., Momoh, M., Fullah, M., Dudas, G., et al. (2014). Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science *345*, 1369–1372.

Ho, S.Y.W., Phillips, M.J., Cooper, A., and Drummond, A.J. (2005). Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol. Biol. Evol. *22*, 1561–1568.

Hoenen, T., Safronetz, D., Groseth, A., Wollenberg, K.R., Koita, O.A., Diarra, B., Fall, I.S., Haidara, F.C., Diallo, F., Sanogo, M., et al. (2015). Virology. Mutation rate and genotype variation of Ebola virus from Mali case sequences. Science *348*, 117–119.

Köster, J., and Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. Bioinformatics *28*, 2520–2522.

Kugelman, J.R., Sanchez-Lockhart, M., Andersen, K.G., Gire, S., Park, D.J., Sealfon, R., Lin, A.E., Wohl, S., Sabeti, P.C., Kuhn, J.H., and Palacios, G.F. (2015a). Evaluation of the potential impact of Ebola virus genomic drift on the efficacy of sequence-based candidate therapeutics. MBio *6*, e02227–e14.

Kugelman, J.R., Wiley, M.R., Mate, S., Ladner, J.T., Beitzel, B., and Fakoli, L. (2015b). Monitoring of Ebola virus Makona evolution through establishment of advanced genomic capability in Liberia. Emerg. Infect. Dis. *21* http://dx.doi.org/10.3201/eid2107.150522.

Kuhn, J.H., Andersen, K.G., Baize, S., Bào, Y., Bavari, S., Berthet, N., Blinkova, O., Brister, J.R., Clawson, A.N., Fair, J., et al. (2014). Nomenclature- and database-compatible names for the two Ebola virus variants that emerged in Guinea and the Democratic Republic of the Congo in 2014. Viruses *6*, 4760–4799.

Matranga, C.B., Andersen, K.G., Winnicki, S., Busby, M., Gladden, A.D., Tewhey, R., Stremlau, M., Berlin, A., Gire, S.K., England, E., et al. (2014). Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. Genome Biol. *15*, 519.

Murin, C.D., Fusco, M.L., Bornholdt, Z.A., Qiu, X., Olinger, G.G., Zeitlin, L., Kobinger, G.P., Ward, A.B., and Saphire, E.O. (2014). Structures of protective antibodies reveal sites of vulnerability on Ebola virus. Proc. Natl. Acad. Sci. USA *111*, 17182–17187.

Olal, D., Kuehne, A.I., Bale, S., Halfmann, P., Hashiguchi, T., Fusco, M.L., Lee, J.E., King, L.B., Kawaoka, Y., Dye, J.M., Jr., and Saphire, E.O. (2012). Structure of an antibody in complex with its mucin domain linear epitope that is protective against Ebola virus. J. Virol. *86*, 2809–2816.

Park, D., Jungreis, I., Tomkins-Tinch, C., and Lin, M. (2015). viral-ngs: v1.0.0. http://dx.doi.org/10.5281/zenodo.17560.

Pickett, B.E., Sadat, E.L., Zhang, Y., Noronha, J.M., Squires, R.B., Hunt, V., Liu, M., Kumar, S., Zaremba, S., Gu, Z., et al. (2012). ViPR: an open bioinformatics database and analysis resource for virology research. Nucleic Acids Res. *40*, D593–D598.

Pybus, O.G., Rambaut, A., Belshaw, R., Freckleton, R.P., Drummond, A.J., and Holmes, E.C. (2007). Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. Mol. Biol. Evol. *24*, 845–852.

Sanchez, A., Trappier, S.G., Ströher, U., Nichol, S.T., Bowen, M.D., and Feldmann, H. (1998). Variation in the glycoprotein and VP35 genes of Marburg virus strains. Virology *240*, 138–146.

Stadler, T., Kühnert, D., Rasmussen, D.A., and du Plessis, L. (2014). Insights into the early epidemic spread of Ebola in Sierra Leone provided by viral sequence data. PLoS Curr. Published online October 6, 2014. http://dx.doi.org/10.1371/currents.outbreaks.02bc6d927ecee7bbd33532ec8ba6a25f.

Tong, Y.-G., Shi, W.-F., Di Liu, Qian, J., Liang, L., Bo, X.-C., Liu, J., Ren, H.G., Fan, H., Ni, M., et al.; China Mobile Laboratory Testing Team in Sierra Leone (2015). Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Leone. Nature. Published online May 13, 2015. http://dx.doi.org/10.1038/nature14490.

Volz, E., and Pond, S. (2014). Phylodynamic analysis of Ebola virus in the 2014 Sierra Leone Epidemic. PLoS Curr. Published online October 24, 2014. http://dx.doi.org/10.1371/currents.outbreaks.6f7025f1271821d4c815385b08f5f80e.

Wertheim, J.O., and Worobey, M. (2009). Relaxed selection and the evolution of RNA virus mucin-like pathogenicity factors. J. Virol. *83*, 4690–4694.

World Health Organization. (2015). Ebola Situation Reports. http://apps.who.int/ebola/en/ebola-situation-reports.

Yozwiak, N.L., Schaffner, S.F., and Sabeti, P.C. (2015). Data sharing: Make outbreak research open access. Nature *518*, 477–479.

Zahn, R.C., Schelp, I., Utermöhlen, O., and von Laer, D. (2007). A-to-G hypermutation in the genome of lymphocytic choriomeningitis virus. J. Virol. *81*, 457–464.

Supplemental Figures

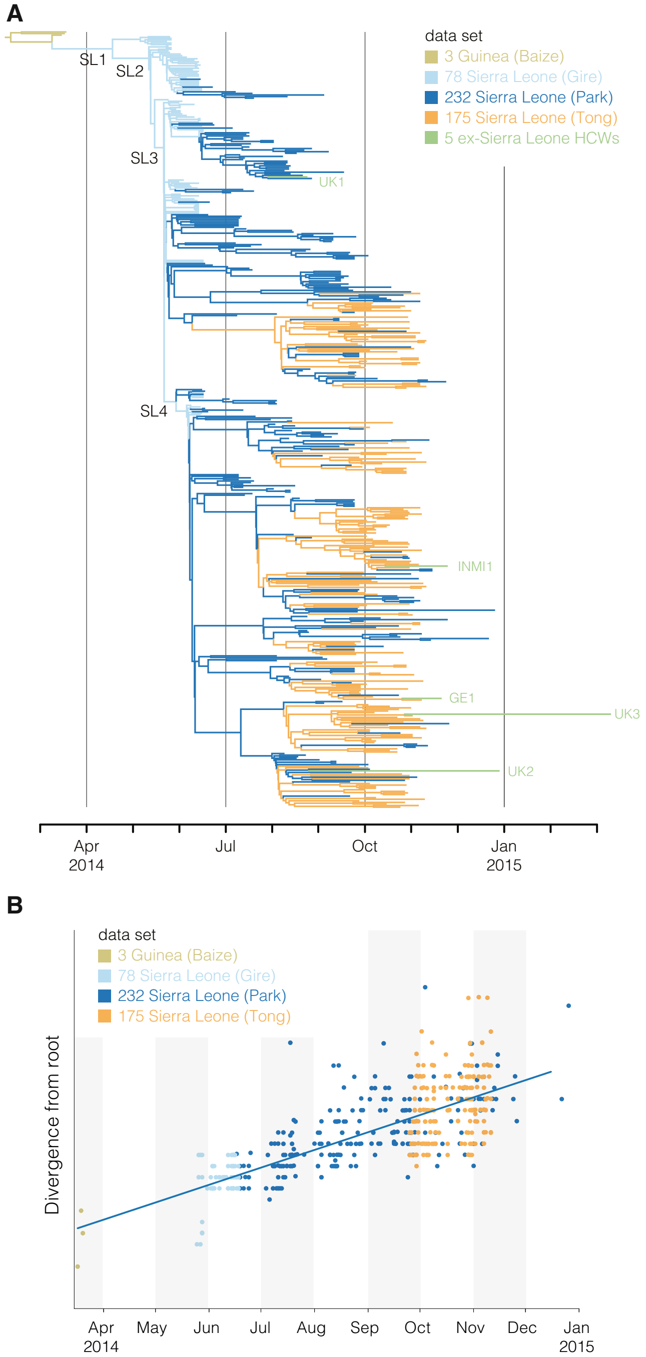


Figure S1. Phylogenetic and Temporal Context of Recent Tong et al. Samples, Related to Figure 1

(A) 175 recently published Ebola virus Makona samples from Sierra Leone (Tong et al., 2015) describe lineages that fall within the genetic diversity of our current dataset (MCC tree from BEAST, as in Figure 1). (B) They span a two month period (Sep 28 to Nov 11, 2014) that falls within the temporal sampling of our current data and shows a consistent evolutionary rate.

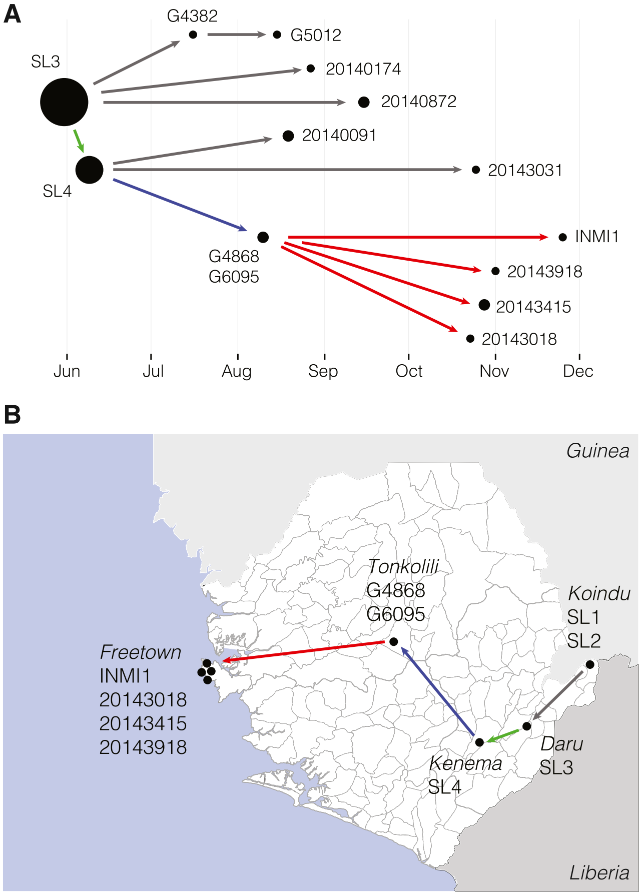


Figure S2. Tracing Historical Ebola Virus Makona Migrations from East to West, Related to Figure 1

(A) Nine Ebola virus (EBOV) Makona genomes (right-hand most circles) from the Freetown area with four groups of apparently ancestral EBOV genomes (middle circles)). Groups of genetically identical genomes (circles) are related to each other by simple vertical relationships (arrows). Solid circles are shown on the date of the earliest sample in the group; the circle area is proportional to the number of samples containing viruses with that genome; arrows represent a set of non-homoplasic SNPs and point from ancestral to derived alleles. Here, “SL3” and “SL4” do not refer to entire clades, but to the viruses that exactly match the canonical SL3 and SL4 genomes with no further mutations. (B) Geographic mapping of one epidemiological route that may account for four of the nine Freetown viruses shown in (A). Groups of identical viruses are shown at their first observed location.

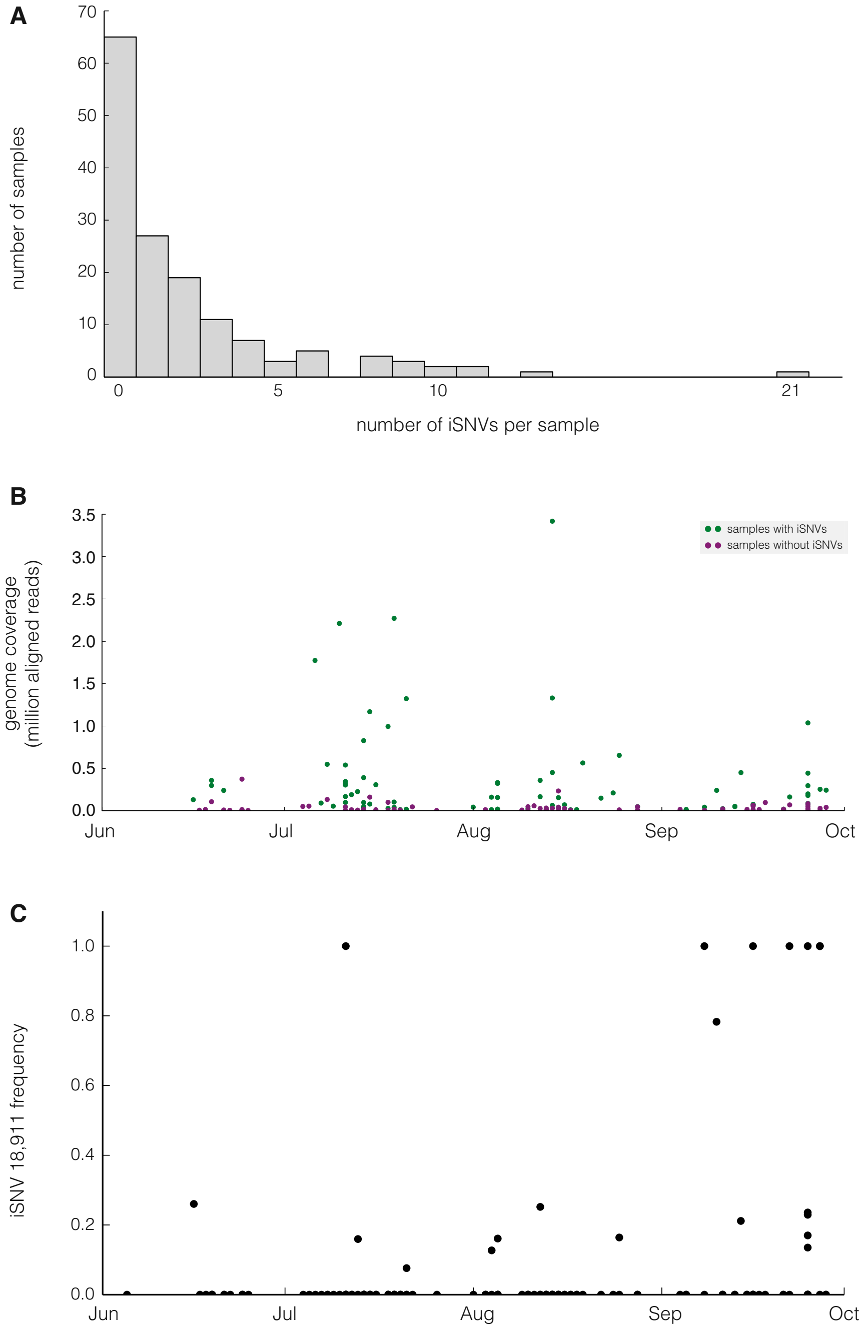


Figure S3. Ebola Virus Makona Intrahost Single-Nucleotide Variants, Related to Figure 2

(A) Distribution of the number of iSNVs per sample. Replicate sequencing and iSNV calling was completed for 150 samples, of which 65 had no iSNV calls. Mean iSNVs per sample (including samples without iSNVs) = 2.04; mean iSNVs per sample (among samples with iSNVs) = 3.6. (B) Sample coverage by date shows the temporal distribution of samples containing Ebola virus (EBOV) genomes with and without iSNV calls. As expected, samples with iSNV calls have generally higher coverage. (C) Intermediate-frequency variants can persist over time with minimal genetic drift, as demonstrated by the iSNV at position 18,911. The existence of intermediate frequency (10%–30%) iSNVs in many different samples over time provides an argument against recurring mutations and may suggest a relatively wide transmission bottleneck between patients.

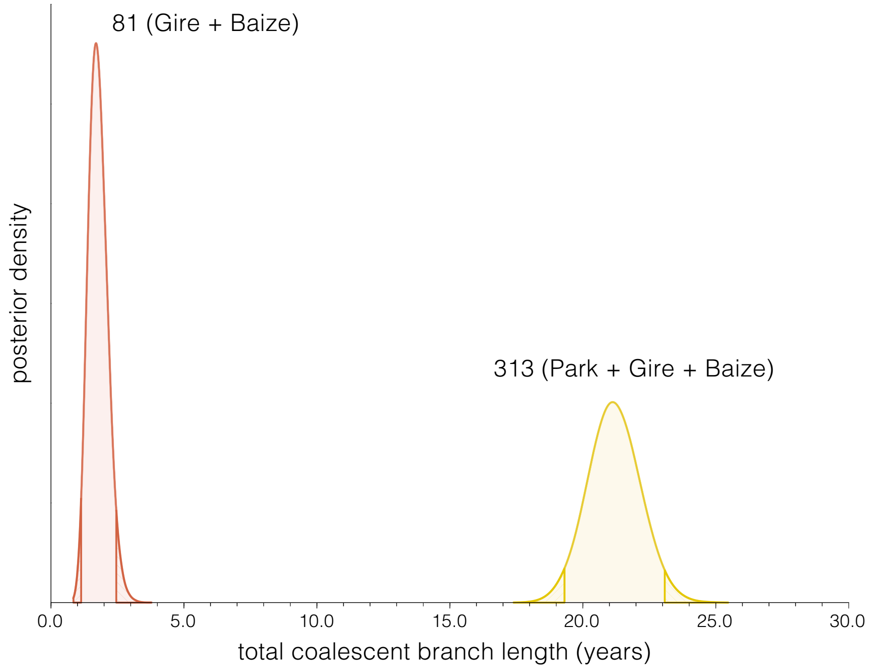


Figure S4. Increased Sampling Improves Evolutionary Rate Estimates, Related to Figure 3

Rate estimates in the recent dataset (Figure 3A) have much tighter credible intervals due to the significantly greater amount of time (total coalescent branch length) compared to the initial outbreak.