



# Evolution and Spread of Ebola Virus in Liberia, 2014–2015

## Citation

Ladner, Jason T., Michael R. Wiley, Suzanne Mate, Gytis Dudas, Karla Prieto, Sean Lovett, Elyse R. Nagle, et al. 2015. Evolution and Spread of Ebola Virus in Liberia, 2014–2015. *Cell Host & Microbe* 18, no. 6: 659–669. doi:10.1016/j.chom.2015.11.008.

## Published Version

doi:10.1016/j.chom.2015.11.008

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:30510307>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)



Published in final edited form as:

*Cell Host Microbe*. 2015 December 9; 18(6): 659–669. doi:10.1016/j.chom.2015.11.008.

## Evolution and spread of Ebola virus in Liberia, 2014–2015

Jason T. Ladner<sup>1,\*,#</sup>, Michael R. Wiley<sup>\*1</sup>, Suzanne Mate<sup>\*1</sup>, Gytis Dudas<sup>2</sup>, Karla Prieto<sup>1</sup>, Sean Lovett<sup>1</sup>, Elyse R. Nagle<sup>1</sup>, Brett Beitzel<sup>1</sup>, Merle L. Gilbert<sup>3</sup>, Lawrence Fakoli<sup>4</sup>, Joseph W. Diclaro II<sup>5</sup>, Randal J. Schoepp<sup>6</sup>, Joseph Fair<sup>7</sup>, Jens H. Kuhn<sup>8</sup>, Lisa E. Hensley<sup>8</sup>, Daniel J. Park<sup>9</sup>, Pardis C. Sabeti<sup>9,10</sup>, Andrew Rambaut<sup>2,11,12</sup>, Mariano Sanchez-Lockhart<sup>1</sup>, Fatorma K. Bolay<sup>4</sup>, Jeffrey R. Kugelman<sup>1,\$</sup>, and Gustavo Palacios<sup>1,\$,#</sup>

<sup>1</sup>Center for Genome Sciences, US Army Medical Research Institute of Infectious Diseases, 1425 Porter Street, Fort Detrick, Frederick, MD, 21702, USA

<sup>2</sup>Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, Charlotte Auerbach Road, Edinburgh EH9 3FL, UK

<sup>3</sup>Molecular and Translational Sciences Division, US Army Medical Research Institute of Infectious Diseases, 1425 Porter Street, Fort Detrick, Frederick, MD, 21702, USA

<sup>4</sup>Liberian Institute for Biomedical Research, Charlesville, Liberia

<sup>5</sup>Naval Medical Research Unit 3, 3A Imtidad Ramses Street, Cairo, Egypt 11517

<sup>6</sup>Diagnostic Systems Division, US Army Medical Research Institute of Infectious Diseases, 1425 Porter Street, Fort Detrick, Frederick, MD, 21702, USA

<sup>7</sup>MRI Global, 1330 Piccard Avenue, Rockville, MD, 20850, USA; The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA

<sup>8</sup>Integrated Research Facility at Fort Detrick, National Institute of Allergy and Infectious Diseases, National Institutes of Health, B-8200 Research Plaza, Fort Detrick, Frederick, MD, 21702, USA

<sup>9</sup>Broad Institute, 75 Ames St, Cambridge, MA, 02142, USA

<sup>10</sup>Harvard University, 52 Oxford Street, Cambridge, MA, 02138, USA

<sup>11</sup>Centre for Immunology, Infection and Evolution, University of Edinburgh, Ashworth Laboratories, Charlotte Auerbach Road, Edinburgh EH9 3FL, UK

<sup>12</sup>Fogarty International Center, National Institutes of Health, 31 Center Drive, Bethesda, MD, 20892, USA

#Correspondence: gustavo.f.palacios.ctr@mail.mil (GP), jason.t.ladner.ctr@mail.mil (JTL).

\*Co-first author

\$Co-senior author

### AUTHOR CONTRIBUTIONS

Conceptualization, JTL, MRW, JRK and GP; Investigation, MRW, SM, KP, ERN, BB, MLG, LF and JRK; Formal Analysis, JTL, MRW, GD, SL and AR; Writing – Original Draft, JTL; Writing – Review & Editing, JTL, MRW, JHK, DJP, MS-L, JRK and GP; Resources, RJS and JF; Supervision, JWD, LEH, PCS, FKB, MS-L and GP; Funding Acquisition: GP

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## SUMMARY

The 2013–present Western African Ebola virus disease (EVD) outbreak is the largest ever recorded with >28,000 reported cases. Ebola virus (EBOV) genome sequencing has played an important role throughout this outbreak; however, relatively few sequences have been determined from patients in Liberia, the second worst-affected country. Here, we report 140 EBOV genome sequences from the second wave of the Liberian outbreak and analyze them in combination with 782 previously published sequences from throughout the Western African outbreak. While multiple early introductions of EBOV to Liberia are evident, the majority of Liberian EVD cases are consistent with a single introduction, followed by spread and diversification within the country. Movement of the virus within Liberia was widespread and reintroductions from Liberia served as an important source for the continuation of the already ongoing EVD outbreak in Guinea. Overall, little evidence was found for incremental adaptation of EBOV to the human host.

## INTRODUCTION

Late in 2013, a novel variant of Ebola virus [EBOV Makona (Kuhn et al., 2014)] emerged in the human population of southeastern Guinea to start what would become the largest human Ebola virus disease (EVD) outbreak on record (Briand et al., 2014). As of 30 September 2015, 28,424 EVD cases had been reported in association with this outbreak, including 11,311 deaths (39.8% case-fatality rate), and new EVD cases were still being reported in two of the three most heavily impacted countries: Guinea (3,805 cases, 13.4% of total) and Sierra Leone (13,911, 48.9%) (WHO, 2015). The unprecedented scale of this outbreak resulted in sustained human-to-human transmission, the ramifications of which are still being explored.

High-throughput EBOV genome sequencing has served an integral role in understanding and responding to the Western African EVD outbreak. As of 30 September 2015, almost 1,000 nearly full-length EBOV Makona genome sequences have been determined (Baize et al., 2014; Bell et al., 2015; Carroll et al., 2015; Castilletti et al., 2015; Gire et al., 2014; Hoenen et al., 2014; Kugelman et al., 2015b; Park et al., 2015; Simon-Loriere et al., 2015; Tong et al., 2015). Early sequences from patients in Guinea and Sierra Leone demonstrated that the Western African outbreak resulted from a single EBOV introduction event followed by sustained human-to-human transmission (Baize et al., 2014; Gire et al., 2014); real-time genomic surveillance in Liberia has provided evidence for sexual transmission of the virus, resulting in changes to public-health policy (Christie et al., 2015; Mate et al., 2015); and continued genomic sequencing throughout the outbreak has provided a detailed view into the ongoing spread and diversification of EBOV, thus providing critical information for maintaining effective control strategies (Carroll et al., 2015; Kugelman et al., 2015a; Park et al., 2015; Simon-Loriere et al., 2015; Tong et al., 2015).

Despite the multi-lateral sequencing efforts, the available genome sequences still represent <4% of the reported cases. Lacking in particular are sequences from cases in Liberia (deposited genomes represent <0.5% of reported cases). As of 30 September 2015, the second highest number of EVD cases among all affected countries (10,672, 37.5% of total) and the highest number of EVD-related deaths (4,808, 42.5% of total) were reported from

Liberia. The Liberian portion of the outbreak is thought to have consisted of at least two distinct waves of EVD cases. The first began in March 2014 and is thought to have been relatively short in duration and scope (<20 reported cases), while the second wave, which likely started during May 2014, included the vast majority of cases (Arwady et al., 2015). Two additional, isolated EVD case clusters have occurred in Liberia, 20–28 March and 28 June – 22 July 2015. However, current data indicate that both likely represent reemergence of transmission chains from the major second wave of EVD cases (Mate et al., 2015) (unpublished data). Liberia was declared free of EBOV infections, for the second time, on 3 September 2015.

Here, we report 140 EBOV genome sequences from the second wave of the Liberian outbreak; these sequences were generated as part of our ongoing surveillance efforts (Kugelman et al., 2015b). Combined with previously published data, these sequences cover 13 Liberian counties and enabled us to perform a longitudinal analysis spanning nearly an entire year of the outbreak. This analysis provides an in-depth look at the introduction and spread of EBOV in Liberia. We further analyzed all Liberian sequences in combination with 734 genomes from Guinea, Mali and Sierra Leone to place the Liberian cases in the context of the entire Western African outbreak and to screen for patterns of EBOV adaptation to the human host.

## RESULTS

### Ebola virus genomes from Liberia

Using high-throughput sequencing technologies, we assembled an additional 140 EBOV genomes from 139 Liberian EVD patients. Two different samples were sequenced from one patient; these had identical consensus sequences. We also improved genome coverage for most of the 25 Liberian EBOV sequences we reported previously (additional 11% on average) (Kugelman et al., 2015b). Together, these 165 genomes spanned 23 June 2014 – 14 February 2015. Genome coverage ranged between 67.6–99.7% with a mean coverage of 98.5% (Table S1); 115 genomes are coding-complete, the remainder are standard drafts according to the nomenclature laid out in (Ladner et al., 2014). In combination with 22 sequences reported from the European Mobile laboratory (1 April – 22 Aug 2014) (Carroll et al., 2015) and one sequence reported from the Centers for Disease Control and Prevention (3 Aug 2014) (Albariño et al., 2015), we analyzed a total of 188 Liberian EBOV genomes. Together, these samples represent ~1.8% of the reported EVD cases in Liberia (as of 30 September 2015) (WHO, 2015). They temporally span nearly one year of the epidemic, including the period during which 99% of the confirmed and probable cases were reported in Liberia (Figure 1). County of origin was reported for 119 (63%) samples. Together they covered 13 of the 15 Liberian counties with sample sizes per county roughly proportional to the number of reported cases.

To place these Liberian sequences within the broader context of the Western African EVD outbreak, we compared them to 734 published sequences from Guinea, Mali and Sierra Leone (Table S2), which together spanned 17 March 2014 – 31 January 2015 (Baize et al., 2014; Carroll et al., 2015; Gire et al., 2014; Hoenen et al., 2015; Park et al., 2015; Simon-Loriere et al., 2015; Tong et al., 2015). Combined, these 922 samples contained 1,474 single

nucleotide polymorphisms (SNPs), 546 (37%) of which were parsimony-informative and 13 insertions/deletions (indels), 8 (61.5%) of which were parsimony-informative. All of the indels were located within non-coding regions. Of the SNPs, 960 were located within coding regions, 403 were non-synonymous within at least one open reading frame (ORF), 555 were synonymous and two were nonsense mutations. In total, three mutations were predicted to result in premature stop codons within at least one ORF [one each in the nucleoprotein (NP), the glycoprotein (GP) and the viral protein 30 (VP30) genes] and one in the loss of the stop codon at the end of the ORF encoding the small soluble glycoprotein (ssGP) in the GP gene. However, with the exception of the previously reported nonsense mutation in VP30 that was present in 29 genomes in our analysis (position 9,354 relative to Ebola virus/H.sapiens-wt/GIN/2014/Makona-C15, GenBank #KJ660346.2) (Kugelman et al., 2015b), all of these mutations were only observed in a single sample obtained from a public repository; therefore, we were unable to verify these putative ORF disruptions.

### Multiple introductions of Ebola virus to Liberia

Phylogenetic analysis revealed the presence of at least three distinct lineages of EBOV in Liberia (Figure 2). The earliest Liberian EBOV genome sequence (1 April 2014), and the only sequence from the first wave of Liberian EVD cases, was placed within the GN1 lineage (Carroll et al., 2015) [referred to as GUI-1 in (Simon-Loriere et al., 2015)], otherwise exclusively containing sequences from Guinea. The other two EBOV lineages observed in Liberia were both associated with samples collected as early as June 2014, during the second wave of the Liberian outbreak. Four samples (20 June – 3 July 2014) fell within lineage SL1 (Gire et al., 2014), which includes EBOV sequences from Guinea and Sierra Leone. The remaining 183 Liberian sequences, including all of the sequences obtained during this study, fell within lineage SL2 (Gire et al., 2014). While SL2 also includes EBOV sequences from both Sierra Leone and Guinea, the basal haplotype within this lineage has previously only been sampled from Sierra Leone; this same haplotype was detected in two of our Liberian sequences (LIBR10089 and LIBR10237). Through phylogenetic rooting, it has been demonstrated that the SL2 lineage was derived from an SL1 virus (Gire et al., 2014). Four nucleotide substitutions differentiate SL1 and SL2 sequences and thus far no intermediate haplotypes have been uncovered, making it difficult to pinpoint the location of the SL1–SL2 transition. However, the presence of the basal haplotypes from both SL1 and SL2 during May in Sierra Leone suggests that this transition occurred in Sierra Leone, in which case the Liberian SL1 and SL2 sequences represent at least two distinct EBOV introductions into Liberia.

One limitation of the current dataset is the lack of sequences from Lofa County after 22 August 2014. Based on epidemiological data and proximity to the epicenter of the outbreak, Lofa County was one of the most likely entrance points into Liberia for EBOV lineages originating in neighboring countries (Arwady et al., 2015; Sack et al., 2014). Although the number of reported EVD cases in Lofa dropped rapidly starting in September 2014, cases were reported as late as mid-November 2014 (WHO, 2015). Additional introductions into Liberia may have occurred during this period with limited transmission and spread outside of Lofa County.

## Ebola virus diversification and spread within Liberia

Despite multiple early introductions, the vast majority of Liberian EBOV sequences are consistent with a single introduction, most likely of an SL2 virus. The SL2 lineage includes Liberian sequences from as early as 20 June 2014 and all sequences sampled after 3 July 2014. Following the initial introduction of an SL2 EBOV into Liberia, the viral population rapidly diversified within that country, consistent with the exponential increase in EVD cases recorded through August 2014 (Figure 1). The second wave of the Liberian outbreak was dominated by eight sub-lineages of SL2 referred to here as LB1–LB8 (Figure 3). Together these eight sub-lineages contained ~92% (167/182) of the Liberian samples stemming from the SL2-type virus introduction. Each Liberian sub-lineage can be distinguished from the basal SL2 haplotype by 1–2 substitution events (Table S3), and mean posterior estimates for the origin of each of these sub-lineages ranged from 31 May 2014 for LB3 (95% HPD: 9 May – 19 June) to 21 July, 2014 for LB7 (95% HPD: 27 June – 5 Aug) (Table S4). Sample testing dates were used to obtain minimum estimates for the duration of each sub-lineage in Liberia. On average, each sub-lineage circulated for at least 130 days. LB4 viruses exhibited the shortest duration within Liberia, only 46 days, while viruses of three different sub-lineages (LB2, LB3, and LB7) each circulated for >180 days. Several of these sub-lineages are defined by non-synonymous substitutions (Table S3); targeted examinations of these changes will be required to determine whether any led to functional changes.

The Liberian outbreak was characterized by a large amount of within-country movement of EBOV. In spite of missing metadata from 37% of the samples, all of the major sub-lineages were observed in 2 Liberian counties; on average, each sub-lineage was observed in 4.25 counties (Figure 3C). In total, we estimated ~63 instances of between-county exchange of EBOV based on 182 Liberian sequences from samples with SL2 lineage viruses (Figures 4 and S1). The heavily impacted counties near Liberia's capital, Margibi and Montserrado, were the largest exporters of EBOV to the rest of the country followed by Lofa, which shares a border with both Guinea and Sierra Leone and is the nearest Liberian county to the putative index case of the Western African outbreak (Méliandou, Nzérékoré Region, Guinea) (Briand et al., 2014). The high number of predicted Margibi to Montserrado movement events is partly due to the inference of Margibi as the most probable county of origin for the SL2 lineage in Liberia (Figure 4A). This inference should be interpreted cautiously given the limited sampling (relative to the full outbreak) and the large amount of missing metadata associated with early Liberian samples; only 31% (8/26) of the samples from June–July 2014 are associated with county-level metadata (all from Lofa).

## Re-introductions of Ebola virus to Guinea from Liberia

Although we found little evidence for additional movement of EBOV into Liberia following the initial appearance of the SL2 lineage, we obtained evidence for several re-introductions of SL2-derived EBOV from Liberia into Guinea. Viruses belonging to four sub-lineages of the Liberian outbreak were also detected in Guinea (LB1, LB2, LB4, and LB5; Figure 3). The Guinean sequences belonging to LB1 and LB4 [GN3 in (Carroll et al., 2015)] each formed distinct clades (for defining SNPs see Table S3), suggesting single introductions of viruses of each sub-lineage followed by spread and diversification within Guinea. Viruses

belonging to LB4 seem to have disappeared from Liberia shortly after spreading into Guinea, where they continued to circulate for ~6 months (Figure 3B). Viruses belonging to LB2 included two divergent Guinean sequences, suggesting two independent introductions each of which resulted in limited, if any, spread within Guinea. Viruses belonging to LB5 [GN4 in (Carroll et al., 2015), GUI-2 in (Simon-Lorieri et al., 2015)] were the most successful of the exports into Guinea both in terms of the number of EVD cases and geographic breadth. In total, 68 of the Guinean EBOV genomes belonged to LB5, including 23 from samples collected in western Guinea (Conakry, Dubréka, Coyah and Forécariah). The other Liberian sub-lineages that were introduced to Guinea were restricted in our dataset to the eastern half of Guinea. The pattern of shared diversity between Guinea and Liberia within the LB5 sub-lineage suggests either multiple EBOV importations from Liberia into Guinea, or movement back into Liberia from Guinea following the initial introduction.

The four sequences from Mali also fell within the LB5 sub-lineage. Their placement is consistent with previous genetic characterization and epidemiological reports (Hoenen et al., 2015; Kugelman et al., 2015b). These four sequences include representatives from two independent introductions of EBOV to Mali, both of which have been traced to the movement of infected individuals from Guinea (Hoenen et al., 2015). Our genetic data additionally demonstrate that both of these introduction events fall within transmission chains that can be traced back to the Liberian portion of the outbreak.

We estimated that viruses from LB1, LB4 and LB5 first entered Guinea in June–July 2014. The time to the most recent common ancestor (TMRCA) for all Guinean sequences in each of these sub-lineages was estimated to be 22 July, 9 July and 22 June for LB1, LB4 and LB5, respectively (Table S5). Circulation of LB1 viruses appears to have subsided around the same time in Liberia and Guinea, in late September 2014. LB4 and LB5 viruses, on the other hand, continued to circulate in Guinea into January 2015, beyond the last sampled genomes from those sub-lineages in Liberia (Figure 3B).

### **Ebola virus evolution in Western Africa**

The ongoing outbreak of EVD in Western Africa is substantially larger than all previously recorded outbreaks combined in terms of both the number of cases and the timespan (a proxy for the length of transmission chains). Therefore, this outbreak has provided EBOV with unprecedented opportunity to evolve within the human host. The non-synonymous substitutions that have arisen during this outbreak are non-uniformly distributed across the EBOV genome (Figure 5A), with peaks in density at the C-terminus of the NP gene, the region of the GP gene encoding the mucin-like domain, and towards the end of the L (RNA-dependent RNA polymerase) gene. Relative frequencies of non-synonymous substitutions across the EBOV genome within the human population are generally consistent with patterns of EBOV divergence within its unknown reservoir host, estimated using the earliest available genome sequence from each of nine distinct EVD outbreaks (Figure 5A; linear  $R^2$  of 0.72). However, the rate of non-synonymous substitutions, relative to synonymous substitutions, is generally elevated across the genome within the Western African outbreak (Figure 5), consistent with previous reports of incomplete purifying selection (Gire et al., 2014; Park et al., 2015; Simon-Lorieri et al., 2015). Peaks in non-synonymous divergence

largely correspond to regions predicted to be intrinsically unstructured (Olabode et al., 2015). Therefore, these hotspots for non-synonymous substitution have likely resulted primarily from a lower level of functional constraint on encoded viral proteins.

We utilized a Bayesian, phylogeny-based approach to identify signatures of diversifying positive selection. This analysis highlighted 11 codons with significant support for positive selection ( $dN>dS$  posterior probability  $>0.9$ ; Figure 5, Table S6). One codon was identified within each of four ORFs: NP (codon 737), VP30 (codon 248), VP24 (codon 28) and L (codon 1,772); the other seven codons were located in the GP<sub>1,2</sub> ORF (codons 82, 455, 472, 479, 480, 493 and 638). Five codons each exhibited two distinct non-synonymous substitutions; the rest contained a single non-synonymous change. The frequency of the non-synonymous variants ranged from 0.11% to 91.7%. Eleven (68.8%) of the sixteen non-synonymous substitutions at these codons were each only observed in the genome from one sample. Five of the significant GP codons and the one L codon are located within regions we identified to be intrinsically unstructured and to be hotspots for non-synonymous substitution. The most significant codon, GP<sub>1,2</sub>-82 ( $dN>dS$  posterior probability = 0.995), includes a non-synonymous substitution (A82V) that arose early in the outbreak, during the transition between the GN1 and SL1 lineages. Codon 82 is located in the region of the GP<sub>1</sub> that contains the receptor binding domain [codons 54–201 (Kuhn et al., 2006)]. The substitutions at GP<sub>2</sub>-638 (Q638R or Q638L) are also intriguing because they affect the glycoprotein's tumor necrosis factor- $\alpha$  converting enzyme (TACE) cleavage site. The sheddase TACE cleaves the membrane-bound GP at the sequence L635-P-D↓-Q removing GP<sub>1,2</sub> from the cell surface. TACE cleavage has been proposed as a mechanism of pathogenesis for filoviruses and sequence conservation among filoviral groups suggests that TACE-cleavage is important for fitness (Dolnik et al., 2004).

## DISCUSSION

Genomic analysis of samples collected over 11 months of the 2013–present Western African EVD outbreak has provided a detailed view into the introduction and spread of EBOV in Liberia. Porous national borders are thought to have played an important role in the spread and maintenance of the ongoing EVD outbreak (Richards et al., 2015; Sack et al., 2014); however, despite multiple early EBOV introductions from Guinea and/or Sierra Leone, the majority of Liberian EVD cases are consistent with a single introduction, likely of an SL2 lineage virus. While infected individuals may have continued to enter Liberia from neighboring countries, these transmission chains did not substantially contribute to the Liberian portion of the outbreak. Several broadly focused measures were taken by the Liberian government to minimize and contain the movement of EBOV into Liberia. These include the early dispatch of response teams to Liberian counties bordering EBOV-infected countries (UNICEF, 2014), an aggressive campaign focused on public outreach and education and the official closure of most of the country's border crossings (MacDougall et al., 2014).

The first appearance of EBOV in Liberia involved a GN1 lineage virus. GN1 sequences were commonly found in eastern Guinea during March–May 2014 (Carroll et al., 2015), consistent with a Guinean source for the first wave of Liberian EVD cases, which began in



mid-March 2014 and ended in early-April 2014 (Arwady et al., 2015). This initial wave of cases is thought to have been locally contained, and our analysis is consistent with this speculation as no second wave Liberian sequences clustered within the GN1 lineage. Unfortunately, county-level information was not associated with the only first wave Liberian sample included in our analysis (KR817194). However, given the timing, it is likely that this sample was part of the transmission chain that began with a woman who died in mid- to late-March 2014 in Lofa County, which is situated in northwestern Liberia and shares borders with both Sierra Leone and Guinea. Six cases from this transmission chain were confirmed to be EBOV positive, including four samples from Lofa and two from Margibi (UNICEF, 2014).

Contact tracing has revealed at least three potential introductions of EBOV to Liberia from Sierra Leone in late-May–early-June 2014. The timing of these events is consistent with the start of the second wave of Liberian EVD cases, and both lineages we observed in this second wave (SL1 and SL2) were present in Sierra Leone at this time (Gire et al., 2014). This time period is also consistent with our previous estimate for the TMRCA of all Liberian SL2 viruses (Kugelman et al., 2015b). The earliest documented introduction involved a patient who traveled to Lofa from Sierra Leone on 23 May 2014; she died in Lofa on 25 May and her body was returned to Sierra Leone for burial (UNICEF, 2014). This patient has been linked to additional EVD cases in Sierra Leone and Liberia, including cases in Monrovia (Sack et al., 2014). However, our analysis indicated that this introduction involved an SL1 virus, which means that it is unlikely to have led to the majority of Liberia’s EVD cases. Two of the sequences analyzed here came from samples that can be linked to this transmission chain (KR817231 & KR817233) (S. Fink, personal communication), and both belong to the SL1 lineage. A second introduction from Sierra Leone to Lofa occurred in early-June 2014 (Sack et al., 2014); it is unclear what lineage of virus was introduced in this instance or whether this case resulted in further transmission in Liberia.

The third documented introduction is reported to have occurred in late-May or early-June 2014 when a woman traveled from Sierra Leone to the New Kru Town community in Monrovia, Montserrado County. This patient has been linked to several other EVD cases in Monrovia, including health care workers at Redemption Hospital (UNICEF, 2014). We don’t know the lineage of EBOV involved in this introduction, but this is a good candidate for the SL2 introduction that appears to have led to the majority of Liberian EVD cases. Our analysis indicated that the success of SL2 lineage viruses in Liberia was likely due in part to the establishment of this lineage in high-density neighborhoods around Monrovia (Montserrado and Margibi Counties). Approximately 70% of Liberia’s reported cases occurred in this region, and it served as an important source of infections in other parts of the country. Additionally, New Kru Town was one of two primary epicenters of EVD cases (in addition to Foya, Lofa County) during June–early-July 2014, at the beginning of the second wave of Liberian EVD cases (UNICEF, 2014).

As a result of demographic transition and internal conflict, Western Africa has become a region characterized by high rates of migration (Gnisci et al., 2006), and this movement is reflected in the spread of EBOV within Liberia. In the 2008 Liberian census, 54% of the

population over the age of 14 reported being internally displaced (Alexander et al., 2015), and 22% of the Liberian-born population was enumerated in a county different from that of birth (LISGIS, 2009). This high rate of migration has resulted in the establishment of strong social ties across geographic regions, and the relatively small size of the country (~111,000 km<sup>2</sup>) makes regular travel between many regions feasible. Reflective of this aspect of Liberian society, we saw widespread movement of EBOV within Liberia, which is likely to have played an important role in the magnitude and longevity of the Liberian portion of the EVD outbreak. Regular migration of infected individuals complicates surveillance and isolation efforts, which are critical for controlling EVD outbreaks (Lindblade et al., 2015; Pandey et al., 2014).

The dominant pathways of EBOV spread within Liberia are broadly consistent with expectations based on the distribution of EVD cases and available data on contact tracing. Our analysis identified two neighboring counties, Montserrado and Margibi, as the primary sources for the spread of EBOV to other Liberian counties. This finding is consistent with epidemiological investigations into EVD clusters in remote Liberian villages. Together, these two counties were identified as the sources for 90% (9/10) of the EVD case clusters for which an index case was successfully identified (Kateh et al., 2015). Montserrado and Margibi were also the two worst-affected Liberian counties, respectively, in terms of the number of reported EVD cases. Similarly, Lofa was identified as the third most important source of EBOV within Liberia, which is consistent with Lofa being the third worst-affected county and a major epicenter early in the second wave of Liberian EVD cases (UNICEF, 2014). However, based on our current dataset, Lofa's contribution as an EBOV source is substantially lower than that of Montserrado and Margibi. This pattern is probably reflective of Lofa's remote location. Lofa is a likely entry point for EBOV to Liberia due to its proximity to the putative origin of the Western African EVD outbreak and shared borders with Guinea and Sierra Leone. However, Lofa is largely isolated from the highly populated regions of Liberia due to poor connecting roads. Therefore, the contribution of human movement to and from Lofa was likely overshadowed by more frequently traveled routes once EBOV became established in the more densely populated counties of Montserrado and Margibi. Our genomic analysis also identified several connections between counties that are consistent with documented, but relatively uncommon movement events (Figure S1), thus illustrating the utility of genomic sequencing to identify and confirm chains of transmission in the absence of good epidemiological data.

The EBOV transmission pattern we deduced for Liberia, driven primarily by within-country spread and diversification, is very similar to that described for eastern Sierra Leone during May 2014–January 2015 (Park et al., 2015), but distinct from the developing picture of the Guinean portion of the outbreak, which appears to have included multiple re-introductions of EBOV from both Liberia and Sierra Leone (Simon-Loriere et al., 2015). This difference in transmission dynamics may partly explain differences between countries in the distribution of EVD cases over time. The portions of the outbreak in Liberia and Sierra Leone both exhibited a single primary peak in cases, whereas the Guinean portion of the outbreak has been characterized by several distinct peaks of similar magnitude (WHO, 2015). Our combined analysis of genomic data from samples collected in four Western African countries demonstrated that occasional importation of EBOV from Liberia likely

played a role in the continuation of the Guinean outbreak and the spread of EBOV to Mali. Starting in August 2014, 70% (69/99) of sequences from the eastern half of Guinea and 30% of sequences from the western half of Guinea (23/76) belonged to evolutionary sub-lineages that originated in Liberia (Figure 6). At least five distinct transmission events from Liberia into Guinea are supported by our analysis, and at least three of these led to sustained EBOV transmission within Guinea. One of these imported sub-lineages, LB5, was further transmitted, on two separate occasions, from Guinea to Mali (Hoenen et al., 2014). We were able to place the ancestors of the three most successful imported lineages within June–July 2014, which is before the official border closings and just before the largest of the peaks in Guinean EVD cases (Figure 6).

It is important to note that international movement of EBOV is only visible in our analysis when this movement resulted in further transmission of the virus. Therefore, we are unable to determine whether the unidirectionality of the international exchange we detected is the result of differences in rates of human movement into/out of Liberia or whether this reflects discrepancies in detecting and controlling newly introduced EBOV transmission chains. The rapid establishment of treatment and isolation facilities was shown to have been effective for interrupting EBOV transmission in several isolated portions of the Liberian outbreak (Lindblade et al., 2015). A detailed investigation of EBOV control measures throughout Western Africa, in light of the movement patterns highlighted in our analysis, will be illustrative regarding the effectiveness of different management approaches.

During “species jumps”, viruses are expected to experience a low-fitness valley, the “depth” of which will vary depending on the virus and hosts involved (Parrish et al., 2008). The absence of EBOV sequences from the unknown reservoir and from infected humans during the first few months of the outbreak (Dec 2013–Feb 2014) prevents the identification of the initial mutations that occurred during the transition of the EBOV Makona variant to humans. However, the large number of EBOV sequences available from the remainder of the outbreak allows for an investigation into patterns of incremental (human) host adaptation following the initial jump. As a whole, we are seeing little evidence for additional adaptation of the virus to humans following the initial transition. We identified 11 codons with evidence of diversifying positive selection. However, viruses with substitutions in these codons were generally rare within the analyzed dataset, suggesting that they are unlikely to have significantly contributed to the magnitude of the Western African EVD outbreak. Furthermore, general patterns of divergence within the Western African outbreak are consistent with divergence within the reservoir host. This result is consistent with other reports on EBOV evolution within Sierra Leone (Gire et al., 2014; Olabode et al., 2015; Park et al., 2015). The lack of strong signatures of adaptation may reflect a relatively shallow fitness valley for EBOV between the reservoir host and humans. In this case, ecological factors determining the frequency of interaction between humans and the reservoir may be the dominant factor controlling the frequency of spillover events rather than host-specific fitness determinants. Targeted investigations into the earliest substitutions that occurred during the Western African EVD outbreak will be critical for understanding the transition of EBOV to humans.

## EXPERIMENTAL PROCEDURES

This work was conducted at the Liberian Institute for Biomedical Research (LIBR) as part of the EVD response and EBOV surveillance, and informed consent was not obtained. With the consent of the National Incident Management System of the Ebola Virus Disease Outbreak and the Liberian Ministry of Health and Social Welfare, the work was supervised by the LIBR institutional review board. All the information obtained from the participants was anonymized for this report.

### Genome Sequencing

In total, we processed 399 real-time PCR-confirmed EBOV-positive samples (Ct values 15.5–26.5) from the Liberian portion of the 2013–present Western African EVD outbreak that had been banked at the Liberian Institute for Biomedical Research; 165 of these (from 164 patients) yielded high-quality consensus genomes. Viral RNA was extracted and sequenced on the Illumina platform (San Diego, CA, USA) primarily using an unbiased amplification method as previously described (Kugelman et al., 2015b). A subset of the samples with low viral coverage were enriched for EBOV sequences using the TruSeq RNA Access kit (Illumina) modified with probes specific for EBOV prior to sequencing (Mate et al., 2015).

EBOV genomes were assembled by aligning reads to Ebola virus/H.sapiens-wt/SLE/2014/Makona-G3864.1 (KR013754, missing bases in the reference were replaced with consensus calls from complete EBOV genomes); this reference is equivalent to the basal SL2 haplotype (Gire et al., 2014), which is thought to have been ancestral to most of the Liberian transmission chains (Kugelman et al., 2015b). Amplification primers were removed from the sequencing reads using Cutadapt v1.21 (Martin, 2011) and low quality reads/bases were filtered using Prinseq-lite v0.20.4 (-min\_qual\_mean 25 -trim\_left 20 -min\_len 50) (Schmieder and Edwards, 2011). Reads were aligned to the reference using Bowtie2 (Langmead and Salzberg, 2012), duplicates were removed with Picard (broadinstitute.github.io/picard) and a new consensus was generated using a combination of Samtools v0.1.18 (Li et al., 2009) and custom scripts. Only bases with Phred quality score 20 were utilized in consensus calling, and a minimum of 3x read-depth coverage, in support of the consensus, was required to make a call; positions lacking this depth of coverage were treated as missing (i.e., called as ‘N’).

### Phylogenetic Analysis

922 EBOV genomes from Guinea, Mali, Sierra Leone and Liberia were analyzed in BEAST (Drummond et al., 2012) (see Tables S1–S2 for accession numbers). A representative subset of these sequences were also analyzed in PopART v1.7.2 (<http://popart.otago.ac.nz>) using the median-joining haplotype network reconstruction method. For the sake of clarity in visualizing the haplotype network, only sequences with 97% coverage were included, and only 75 sequences were included from each of the primary lineages in Sierra Leone (SL3 and SL4). The genomes included from SL3 and SL4 were chosen randomly from the available sequences with 99.5% genome coverage. This resulted in the inclusion of 175 Liberian genomes and 466 genomes from Guinea, Mali and Sierra Leone (Tables S1–S2).

For BEAST, sites were partitioned into non-coding intergenic regions and codon positions 1, 2 and 3. The evolution of all 4 site partitions was modelled by independent HKY substitution models with gamma(4)-distributed rate heterogeneity with a relaxed molecular clock with lognormally distributed rate categories (Drummond et al., 2006). The non-parametric Bayesian SkyGrid tree prior (Gill et al., 2013) was used. These analyses were used to identify the separate EBOV introductions into Liberia and to select a subset of isolates for more detailed analysis.

Analysis of the entire Western African EBOV sequence dataset revealed one phylogenetic lineage (SL2) that contained the vast majority of Liberian sequences (182/188, 97%). These 182 genomes were used for more detailed analysis, including the inference of within-country migration and the dating of common ancestors. To estimate dates for the movement of EBOV between countries, 97 isolates from Guinea and Mali were added to this Liberian dataset; in the full analysis, these isolates clustered within Liberian SL2 sub-lineages. These analyses used the same model described for the full analysis, but with a prior on the mean of the lognormal distribution ( $N[1.144 \times 10^{-3}, 5.7968 \times 10^{-3}]$ ) informed by the rate estimated for the full Western African sequence dataset. To infer viral migration between counties (within Liberia) an asymmetric continuous time Markov chain (CTMC) approach was chosen. Bayesian stochastic search variable selection (BSSVS) was employed to identify strongly supported (Bayes factor > 3) migrations. An uninformative reference prior was used on the migration rate. The MCMC chain was run twice for 50 million states, with 9,000 samples from each chain combined. For all analyses, the first 10% of the chain (10 million states) was discarded as burn-in. Path-O-Gen version 1.4 (Drummond et al., 2012) was used to calculate the root-to-tip distances by using the “best-fitting root” option and a maximum-likelihood phylogeny [PhyML version 3.0 (Guindon et al., 2010); general time reversible model].

## Molecular Evolution

Sequences were aligned using Sequencher v5.2.3 (Gene Codes, Ann Arbor, MI). Custom scripts were used to generate a variant call file (Data S1), which was then annotated with SnpEff (v4.1b, build 2015-02-13) (Cingolani et al., 2012) using the genome of Ebola virus/H.sapiens-wt/GIN/2014/Makona-C15 (KJ660346.2) as a reference. Nine EBOV genome sequences (1 per outbreak) were used to explore patterns of divergence between spatially and/or temporally distinct outbreaks: NC002549, KC242791, KC242792, KC242796, KC242794, KC242785, HQ613402, KJ660346 and KM519951.

All EBOV genomes with 93% genome coverage (920 genomes) were used to screen for signatures of positive diversifying selection. A maximum-likelihood tree was constructed with PhyML using a GTR model and this tree was used to guide positive selection analysis with FUBAR (Murrell et al., 2013), part of the HyPhy package, using 20 MCMC chains of length 100,000,000 that were sampled 5,000 times each. Codons were predicted to be under positive diversifying selection if the posterior probability that  $dN/dS > 1$  at that site was greater than 0.9.

Intrinsically unstructured regions were predicted using the IUPred webserver with long disorder settings (Dosztanyi et al., 2005). Regions with a disorder tendency 0.4 were considered unstructured.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work was supported by the Defense Threat Reduction Agency, Global Biosurveillance Technology Initiative, Global Emerging Infections System and US Agency for International Development. A.R. was supported by EU Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no. 278433-PREDEMICS. The content of this publication does not necessarily reflect the views or policies of the US Army, the US Dept of Defense, the US Dept of Health and Human Services or the institutions/companies affiliated with the authors. JHK performed this work as an employee of Tunnell Government Services, Inc., a subcontractor to Battelle Memorial Institute under its prime contract with the National Institute of Allergy and Infectious Diseases (Contract No. HHSN2722007000161).

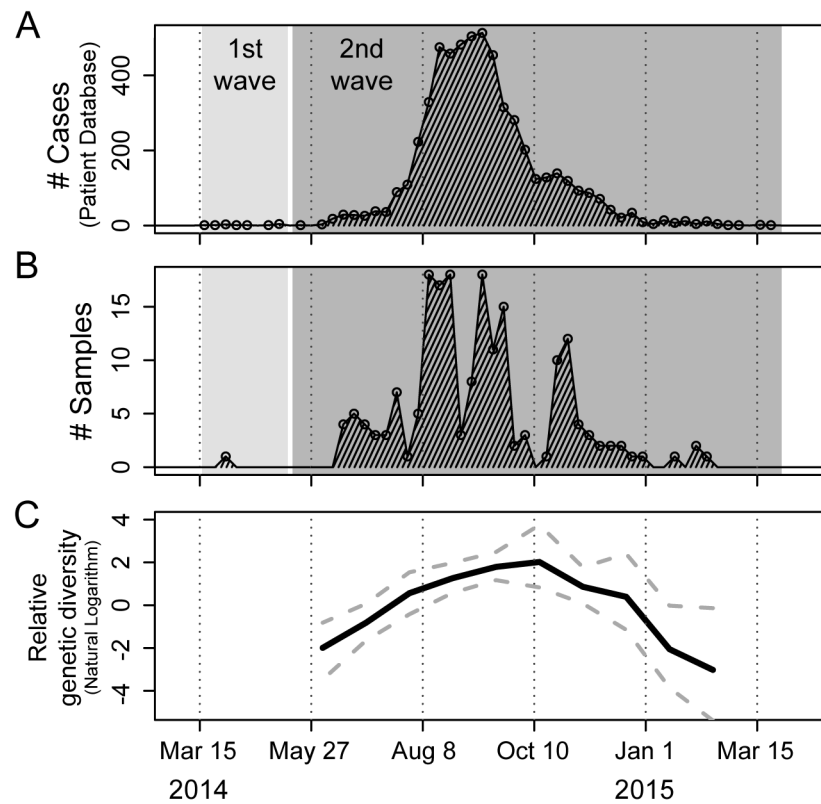
## References

- Albariño CG, Guerrero LW, Lo MK, Nichol ST, Towner JS. Development of a reverse genetics system to generate a recombinant Ebola virus Makona expressing a green fluorescent protein. *Virology*. 2015; 484:259–264. [PubMed: 26122472]
- Alexander KA, Sanderson CE, Marathe M, Lewis BL, Rivers CM, Shaman J, Drake JM, Lofgren E, Dato VM, Eisenberg MC, et al. What Factors Might Have Led to the Emergence of Ebola in West Africa? *PLoS Negl Trop Dis*. 2015; 9:e0003652. [PubMed: 26042592]
- Arwady MA, Bawo L, Hunter JC, Massaquoi M, Matanock A, Dahn B, Ayscue P, Nyenswah T, Forrester JD, Hensley LE, et al. Evolution of Ebola virus disease from exotic infection to global health priority, Liberia, mid-2014. *Emerg Infect Dis*. 2015; 21:578–584. [PubMed: 25811176]
- Baize S, Pannetier D, Oestereich L, Rieger T, Koivogui L, Magassouba N, Soropogui B, Sow MS, Keita S, De Clerck H, et al. Emergence of Zaire Ebola virus disease in Guinea. *The New England journal of medicine*. 2014; 371:1418–1425. [PubMed: 24738640]
- Bell, A.; Lewandowski, K.; Myers, R.; Wooldridge, D.; Aarons, E.; Simpson, A.; Vipond, R.; Jacobs, M.; Gharbia, S.; Zambon, M. Genome sequence analysis of Ebola virus in clinical samples from three British healthcare workers, August 2014 to March 2015. 2015. (<http://www.eurosurveillance.org/images/dynamic/EE/V20N20/art21131.pdf>)
- Briand S, Bertherat E, Cox P, Formenty P, Kieny MP, Myhre JK, Roth C, Shindo N, Dye C. The international Ebola emergency. *The New England journal of medicine*. 2014; 371:1180–1183. [PubMed: 25140855]
- Carroll MW, Matthews DA, Hiscox JA, Elmore MJ, Pollakis G, Rambaut A, Hewson R, Garcia-Dorival I, Bore JA, Koundouno R, et al. Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. *Nature*. 2015
- Castilletti C, Carletti F, Gruber CE, Bordi L, Lalle E, Quartu S, Meschi S, Lapa D, Colavita F, Chiappini R, et al. Molecular Characterization of the First Ebola Virus Isolated in Italy, from a Health Care Worker Repatriated from Sierra Leone. *Genome Announc*. 2015; 3:e00639–00615. [PubMed: 26089420]
- Christie A, Davies-Wayne GJ, Cordier-Lasalle T, Blackley DJ, Laney AS, Williams DE, Shinde SA, Badio M, Lo T, Mate SE, et al. Possible sexual transmission of Ebola virus - Liberia, 2015. *MMWR Morb Mortal Wkly Rep*. 2015; 64:479–481. [PubMed: 25950255]
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*. 2012; 6:80–92. [PubMed: 22728672]

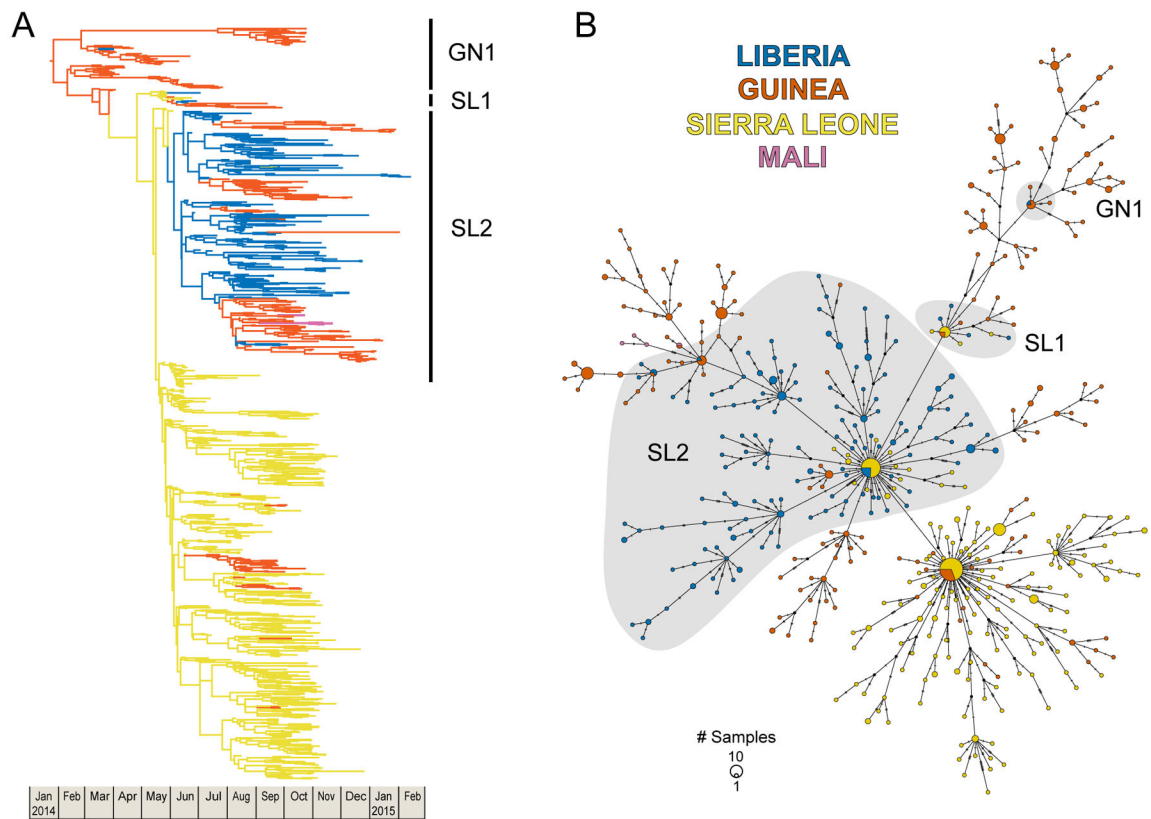
- Dolnik O, Volchkova V, Garten W, Carbonnelle C, Becker S, Kahnt J, Ströher U, Klenk HD, Volchkov V. Ectodomain shedding of the glycoprotein GP of Ebola virus. *The EMBO journal*. 2004; 23:2175–2184. [PubMed: 15103332]
- Dosztanyi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*. 2005; 21:3433–3434. [PubMed: 15955779]
- Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 2006; 4:e88. [PubMed: 16683862]
- Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*. 2012; 29:1969–1973. [PubMed: 22367748]
- Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol Biol Evol*. 2013; 30:713–724. [PubMed: 23180580]
- Gire SK, Goba A, Andersen KG, Sealfon RS, Park DJ, Kanneh L, Jalloh S, Momoh M, Fullah M, Dudas G, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science*. 2014; 345:1369–1372. [PubMed: 25214632]
- Gnisci, D.; Tremolieres, M.; Bossard, L. Robin, N., editor. Atlas on Regional Integration in West Africa. 2006. <http://www.oecd.org/migration/38409521.pdf:ECOWAS-SWAC/OECD>
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010; 59:307–321. [PubMed: 20525638]
- Hoenen T, Groseth A, Feldmann F, Marzi A, Ebihara H, Kobinger G, Gunther S, Feldmann H. Complete genome sequences of three ebola virus isolates from the 2014 outbreak in west Africa. *Genome Announc*. 2014; 2
- Hoenen T, Safronetz D, Groseth A, Wollenberg KR, Koita OA, Diarra B, Fall IS, Haidara FC, Diallo F, Sanogo M, et al. Virology. Mutation rate and genotype variation of Ebola virus from Mali case sequences. *Science*. 2015; 348:117–119. [PubMed: 25814067]
- Kateh F, Nagbe T, Kieta A, Barskey A, Gasasira AN, Driscoll A, Tucker A, Christie A, Karmo B, Scott C, et al. Rapid response to Ebola outbreaks in remote areas - Liberia, July-November 2014. *MMWR Morb Mortal Wkly Rep*. 2015; 64:188–192. [PubMed: 25719682]
- Kugelman JR, Sanchez-Lockhart M, Andersen KG, Gire S, Park DJ, Sealfon R, Lin AE, Wohl S, Sabeti PC, Kuhn JH, et al. Evaluation of the potential impact of Ebola virus genomic drift on the efficacy of sequence-based candidate therapeutics. *mBio*. 2015a; 6
- Kugelman JR, Wiley MR, Mate S, Ladner JT, Beitzel B, Fakoli L, Taweh F, Prieto K, Diclaro JW, Minogue T, et al. Monitoring of Ebola Virus Makona Evolution through Establishment of Advanced Genomic Capability in Liberia. *Emerg Infect Dis*. 2015b; 21:1135–1143. [PubMed: 26079255]
- Kuhn JH, Andersen KG, Baize S, Bao Y, Bavari S, Berthet N, Blinkova O, Brister JR, Clawson AN, Fair J, et al. 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*. 2014; 6:4760–4799. [PubMed: 25421896]
- Kuhn JH, Radoshitzky SR, Guth AC, Warfield KL, Li W, Vincent MJ, Towner JS, Nichol ST, Bavari S, Choe H. Conserved receptor-binding domains of Lake Victoria marburgvirus and Zaire ebolavirus bind a common receptor. *J Biol Chem*. 2006; 281:15951–15958. [PubMed: 16595665]
- Ladner JT, Beitzel B, Chain PS, Davenport MG, Donaldson EF, Frieman M, Kugelman JR, Kuhn JH, O’Rear J, Sabeti PC, et al. Standards for sequencing viral genomes in the era of high-throughput sequencing. *mBio*. 2014; 5:e01360–01314. [PubMed: 24939889]
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9:357–359. [PubMed: 22388286]
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25:2078–2079. [PubMed: 19505943]

- Lindblade K, Kateh F, Nagbe T, Neatherlin J, Pillai S, Attfield K, Dweh E, Barradas D, Williams S, Blackley D. Decreased Ebola Transmission after Rapid Response to Outbreaks in Remote Areas, Liberia, 2014. *Emerging Infect Dis.* 2015; 21:1800. [PubMed: 26402477]
- LISGIS. 2008 Population and Housing Census Final Results. Liberia Institute of Statistics and Geo-Information Services; 2009. [http://www.lisgis.net/pg\\_img/NPHC\\_2008\\_Final\\_Report.pdf](http://www.lisgis.net/pg_img/NPHC_2008_Final_Report.pdf)
- MacDougall, C.; Felix, B.; Craft, D. Liberia shuts border crossings, restricts gatherings to curb Ebola spreading. Reuters; 2014. <http://www.reuters.com/article/2014/07/28/us-health-ebola-africa-idUSKBN0FX00V20140728>
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal.* 2011; 17:10–12.
- Mate SE, Kugelman JR, Nyenswah TG, Ladner JT, Wiley MR, Cordier-Lassalle T, Christie A, Schroth GP, Gross SM, Davies-Wayne GJ, et al. Molecular Evidence of Sexual Transmission of Ebola Virus. *The New England journal of medicine.* 2015
- Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Pond SLK, Scheffler K. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol.* 2013; 30:1196–1205. [PubMed: 23420840]
- Olabode AS, Jiang X, Robertson DL, Lovell SC. Ebolavirus is evolving but not changing: No evidence for functional change in EBOV from 1976 to the 2014 outbreak. *Virology.* 2015; 482:202–207. [PubMed: 25880111]
- Pandey A, Atkins KE, Medlock J, Wenzel N, Townsend JP, Childs JE, Nyenswah TG, Ndeffo-Mbah ML, Galvani AP. Strategies for containing Ebola in West Africa. *Science.* 2014; 346:991–995. [PubMed: 25414312]
- Park DJ, Dudas G, Wohl S, Goba A, Whitmer SL, Andersen KG, Sealfon RS, Ladner JT, Kugelman JR, Matranga CB, et al. Ebola Virus Epidemiology, Transmission, and Evolution during Seven Months in Sierra Leone. *Cell.* 2015; 161:1516–1526. [PubMed: 26091036]
- Parrish CR, Holmes EC, Morens DM, Park EC, Burke DS, Calisher CH, Laughlin CA, Saif LJ, Daszak P. Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol Mol Biol Rev.* 2008; 72:457–470. [PubMed: 18772285]
- Richards P, Amara J, Ferme MC, Kamara P, Mokuwa E, Sheriff AI, Suluku R, Voors M. Social pathways for Ebola virus disease in rural Sierra Leone, and some implications for containment. *PLoS Negl Trop Dis.* 2015; 9:e0003567. [PubMed: 25886400]
- Sack, K.; Fink, S.; Belluck, P.; Nossiter, A. How Ebola Roared Back: Missed Chances. *New York Times*; 2014. [http://www.nytimes.com/2014/12/30/health/how-ebola-roared-back.html?\\_r=0](http://www.nytimes.com/2014/12/30/health/how-ebola-roared-back.html?_r=0)
- Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics.* 2011; 27:863–864. [PubMed: 21278185]
- Simon-Loriere E, Faye O, Faye O, Koivogui L, Magassouba N, Keita S, Thiberge JM, Diancourt L, Bouchier C, Vandebogaert M, et al. Distinct lineages of Ebola virus in Guinea during the 2014 West African epidemic. *Nature.* 2015; 524:102–104. [PubMed: 26106863]
- Tong YG, Shi WF, Di L, Qian J, Liang L, Bo XC, Liu J, Ren HG, Fan H, Ni M, et al. Genetic diversity and evolutionary dynamics of Ebola virus in Sierra Leone. *Nature.* 2015
- UNICEF. UNICEF-Liberia Ebola Virus Outbreak SitReps. 2014. ([www.unicef.org/appeals/files/UNICEF\\_Liberia\\_SitRep11\\_Ebola\\_Virus\\_Outbreak\\_7April2014.pdf](http://www.unicef.org/appeals/files/UNICEF_Liberia_SitRep11_Ebola_Virus_Outbreak_7April2014.pdf), [www.unicef.org/appeals/files/UNICEF\\_Liberia\\_SitRep19\\_Ebola\\_Viral\\_Disease\\_23Apr2014.pdf](http://www.unicef.org/appeals/files/UNICEF_Liberia_SitRep19_Ebola_Viral_Disease_23Apr2014.pdf), [www.unicef.org/appeals/files/UNICEF\\_Liberia\\_SitRep\\_22\\_Ebola\\_Viral\\_Disease\\_2\\_June\\_2014.pdf](http://www.unicef.org/appeals/files/UNICEF_Liberia_SitRep_22_Ebola_Viral_Disease_2_June_2014.pdf), [www.unicef.org/appeals/files/UNICEF\\_Liberia\\_SitRep23\\_Ebola\\_Viral\\_Disease\\_13\\_June\\_2014.pdf](http://www.unicef.org/appeals/files/UNICEF_Liberia_SitRep23_Ebola_Viral_Disease_13_June_2014.pdf), [www.unicef.org/appeals/files/UNICEF\\_Liberia\\_SitRep24\\_Ebola\\_Viral\\_Disease\\_17\\_June\\_2014.pdf](http://www.unicef.org/appeals/files/UNICEF_Liberia_SitRep24_Ebola_Viral_Disease_17_June_2014.pdf), [www.unicef.org/appeals/files/UNICEF\\_Liberia\\_SitRep\\_Ebola\\_Viral\\_Disease\\_9July2014.pdf](http://www.unicef.org/appeals/files/UNICEF_Liberia_SitRep_Ebola_Viral_Disease_9July2014.pdf))
- WHO. Ebola Situation Reports. 2015. (<http://apps.who.int/ebola/ebola-situation-reports>)



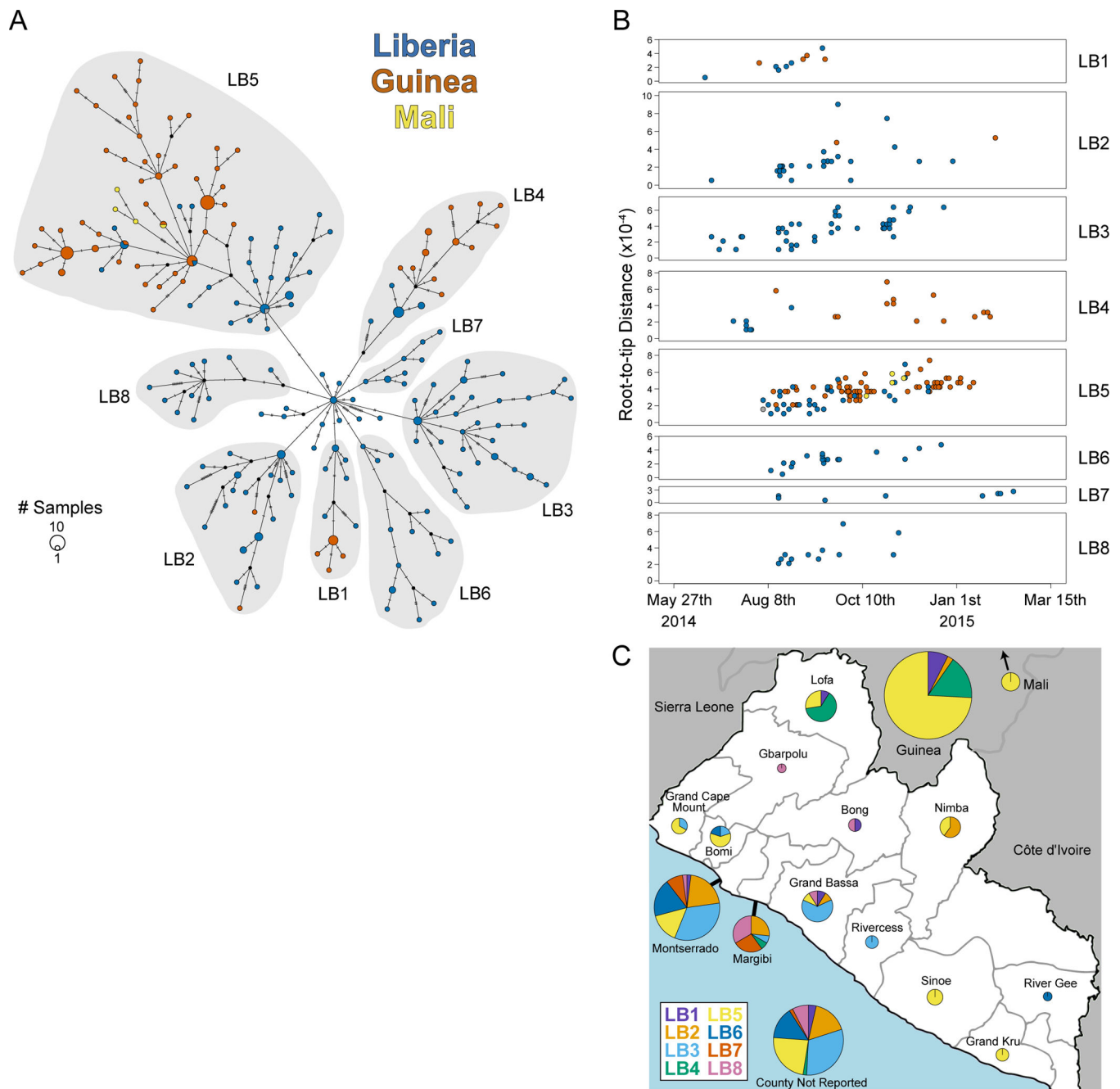


**Figure 1.** Temporal dynamics of the EVD outbreak in Liberia. (A) Confirmed and probable EVD cases in Liberia through time (WHO, 2015). (B) Temporal distribution of the 188 Liberian samples analyzed in this study. (C) Relative genetic diversity calculated with BEAST for the SL2 lineage in Liberia (SkyGrid reconstruction). The solid line represents the median estimate from the posterior probability, and the dashed lines represent the upper and lower estimates of the 95% credible interval.

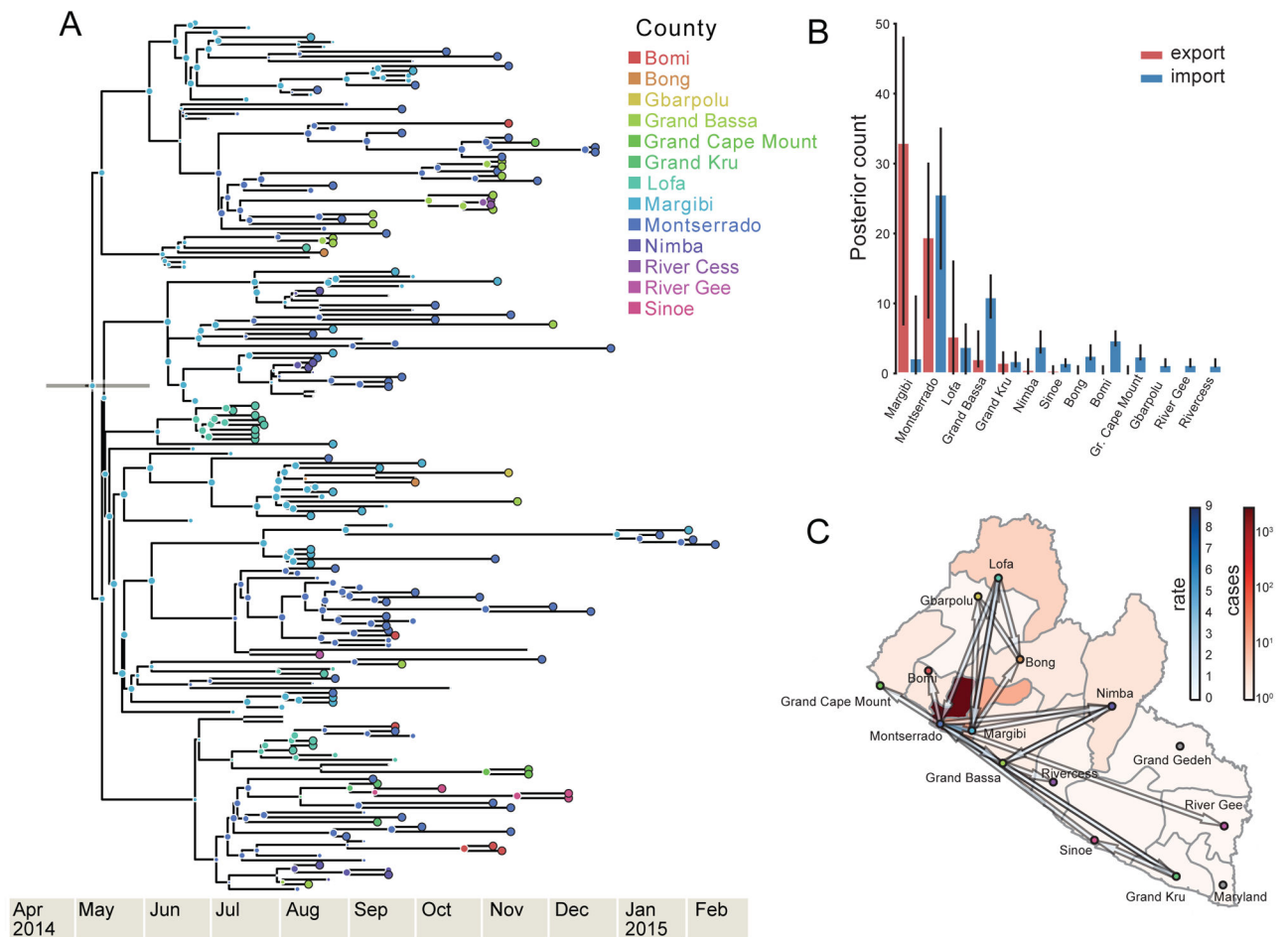


**Figure 2.**

Multiple early introductions of EBOV into Liberia. (A) Phylogenetic and temporal placement of 188 Liberian EBOV genomes relative to 734 EBOV sequences from Guinea, Mali and Sierra Leone. Three distinct lineages are represented in the Liberian samples: GN1, SL1 and SL2. (B) Median-joining haplotype network including 175 Liberian EBOV sequences with 97% genome coverage and 466 EBOV sequences representative of lineages circulating elsewhere in Western Africa.



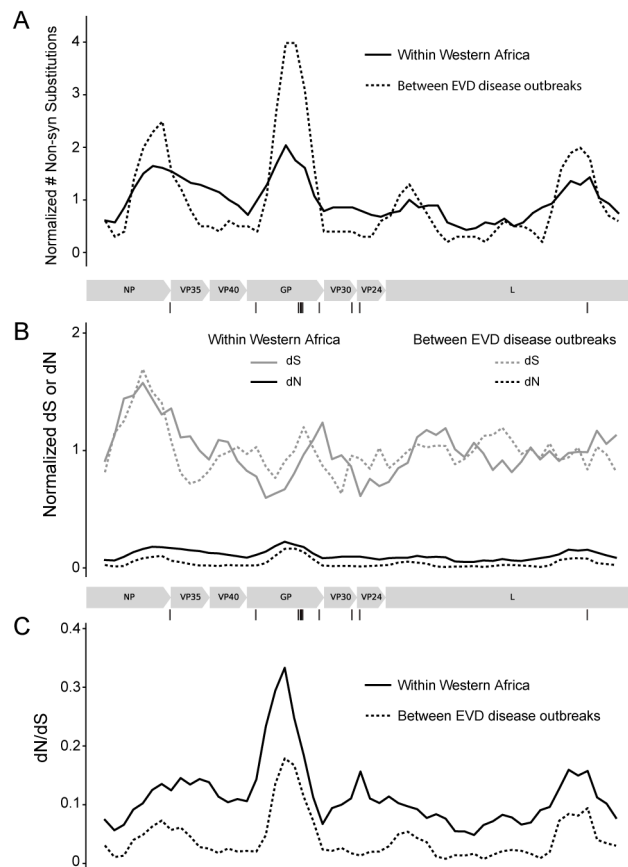
**Figure 3.** Eight primary sub-lineages circulated during the second wave of EVD cases in Liberia. (A) Median-joining haplotype network based on a full genome alignment of 158 sequences from Liberia (SL2 only and with 98% genome coverage) and 95 sequences from Guinea and Mali that clustered within Liberian sub-lineages. (B) Root-to-tip distance versus testing date for each sub-lineage. (C) Geographic distribution of the Liberian sub-lineages, at the county-level within Liberia (white) and country-level outside Liberia (grey).



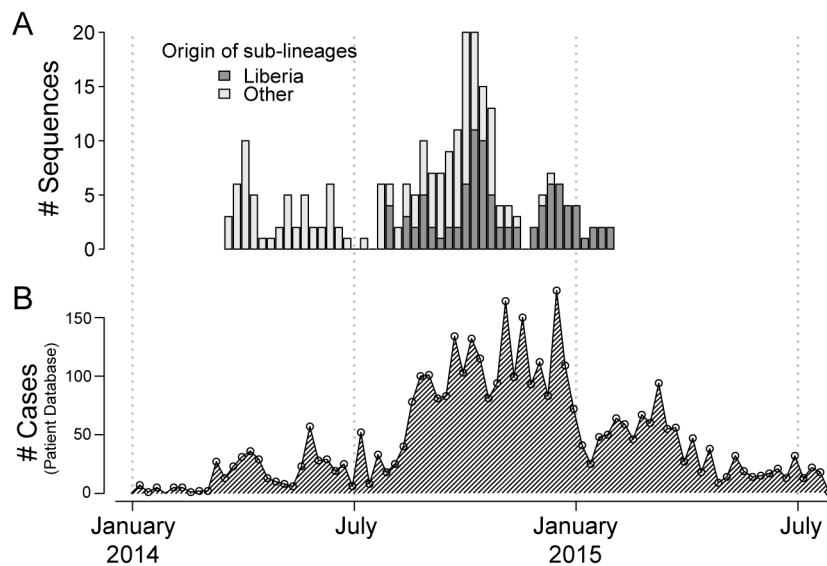
**Figure 4.**

Widespread movement of EBOV during the second wave of the Liberian outbreak. (A) Temporal maximum clade credibility tree from BEAST analysis. Circles at the nodes indicate inferred ancestral location of each lineage. Circles with black outlines at the branch tips represent samples with known county of origin; those with white outlines were inferred in the analysis as a latent variable over the course of the MCMC. Circle size is proportional to the posterior probability of the assigned county. The bar at the root indicates the 95% HPD for the estimated root date. (B) Counts of exported vs imported viral lineages between locations across the posterior distribution. Vertical black lines indicate 95% HPD. (C) Well-supported (Bayes Factor  $\geq 3$ ) asymmetric rates of viral migration between counties. Arrow color indicates magnitude. Counties are colored by cumulative number of cases reported by the WHO.

See also Figure S1.



**Figure 5.** Distribution of synonymous and non-synonymous substitutions within the open reading frames (ORFs) of the EBOV genome. Black lines below the ORFs (grey arrows) indicate the positions of codons with significant evidence of positive selection (Table S6). (A) Non-synonymous substitutions that have occurred within the Western African EVD outbreak (solid line) and between outbreaks caused by EBOV (dashed line). A sliding window of 1000 nucleotides (nt) was used with a step size of 250 nt. Each count was normalized by the average number of substitutions per window. (B & C) Distribution of dS (synonymous substitutions per synonymous site), dN (non-synonymous substitutions per non-synonymous site) and dN/dS. For each dataset, dS and dN were both normalized by the average dS per window. A sliding window of 999 nt (333 codons) was used with a step size of 249 nt (83 codons).



**Figure 6.** Liberian sub-lineages of EBOV contributed substantially to the largest peak in Guinean EVD cases. (A) The number of Guinean EBOV sequences through time colored based on the geographic origin of the evolutionary lineages to which each sequence belongs. (B) Confirmed and probable EVD cases in Guinea through time, according to the WHO's patient database (WHO, 2015).